BARRERAS CONTRA EL SPAM • ASI NEGOCIAN LOS ANIMALES

INVESTIGACION y CIENCIA

JUNIO 2005
6,00 EUROS

Edición española de SCIENTIFIC AMERICAN

LA GEODINAMO

EL OTRO GENOMA

SUPERCONDUCTORES A ALTAS TEMPERATURAS

ORIGEN DE LA MATERIA

CALIDAD OPTICA DEL OJO

EL METODO BOTANICO DE CAVANILLES





Edición española de SCIENTIFIC AMERICAN

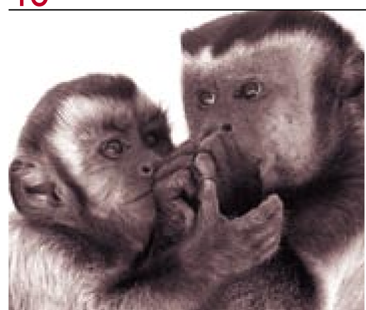
Junio de 2005

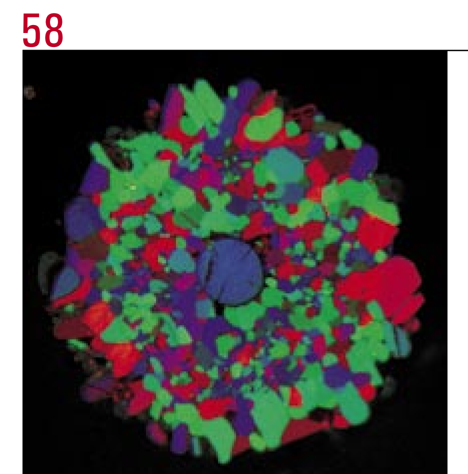
Número 345







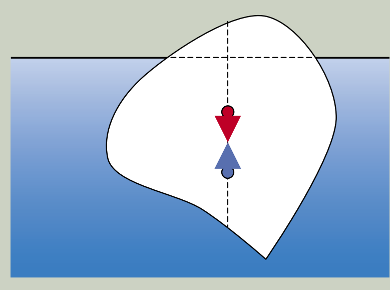


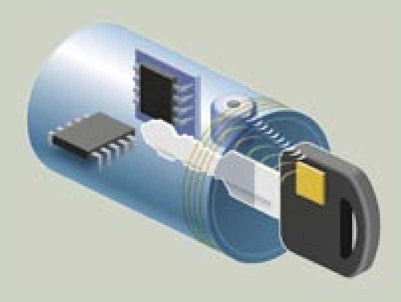


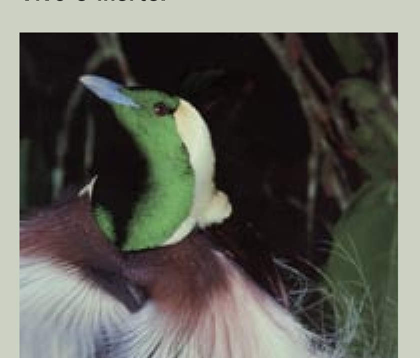














DIRECTOR GENERAL José M.ª Valderas Gallardo
DIRECTORA FINANCIERA Pilar Bronchal Garfella
EDICIONES Juan Pedro Campos Gómez
Laia Torres Casas
PRODUCCIÓN M.ª Cruz Iglesias Capón
Albert Marín Garau
SECRETARÍA Purificación Mayoral Martínez
ADMINISTRACIÓN Victoria Andrés Laiglesia
SUSCRIPCIONES Concepción Orenes Delgado
Olga Blanco Romero
EDITA Prensa Científica, S.A. Muntaner, 339 pral. 1.ª
08021 Barcelona (España)
Teléfono 934 143 344 Telefax 934 145 413
www.investigacionyciencia.es

SCIENTIFIC AMERICAN

EDITOR IN CHIEF John Rennie
EXECUTIVE EDITOR Mariette DiChristina
MANAGING EDITOR Ricki L. Rusting
NEWS EDITOR Philip M. Yam
SPECIAL PROJECTS EDITOR Gary Stix
SENIOR EDITOR Michelle Press
SENIOR WRITER W. Wayt Gibbs
EDITORS Mark Alpert, Steven Ashley, Graham P. Collins, Steve Mirsky, George Musser y Christine Soares
PRODUCTION EDITOR Richard Hunt
GENERAL MANAGER Michael Florek
VICE PRESIDENT AND MANAGING DIRECTOR, INTERNATIONAL Dean Sanderson
PRESIDENT AND CHIEF EXECUTIVE OFFICER Gretchen G. Teichgraeber
CHAIRMAN John Sargent

DISTRIBUCION

para España:

LOGISTA, S. A.
Pol. Ind. Polvoranca
Trigo, 39, Edif. 2
28914 Leganés (Madrid)
Teléfono 914 819 800

para los restantes países:

Prensa Científica, S. A.
Muntaner, 339 pral. 1.ª
08021 Barcelona

PUBLICIDAD

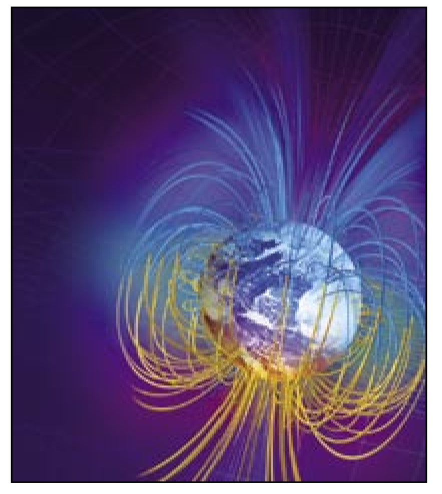
Madrid:
Manuel Martín Martín
Teléfono 670 236 715
e-mail: publicidad-ic@hotmail.com

Cataluña:
QUERALTO COMUNICACION
Julián Queraltó
Sant Antoni M.ª Claret, 281 4.º 3.ª
08041 Barcelona
Tel. y fax 933 524 532
Móvil 629 555 703

COLABORADORES DE ESTE NUMERO

Asesoramiento y traducción:

Sònia Ambrós: *La geodinamo*; Felipe Cortés: *El otro genoma*; Joandomènech Ros: *Así negocian los animales*; Ramón Pascual: *El origen de la materia*; Juan Bartolomé: *El diboruro de magnesio, superconductor a alta temperatura*; Jürgen Goicoechea: *Curiosidades de la física;* J. Vilardell: *Apuntes* e *Ideas aplicadas*.



Portada: Jean-Francois Podevin

SUSCRIPCIONES

Prensa Científica S. A.
Muntaner, 339 pral. 1.ª
08021 Barcelona (España)
Teléfono 934 143 344
Fax 934 145 413

Precios de suscripción:

	Un año	Dos años
España	65,00 euro	120,00 euro
Resto del mundo	90,00 euro	170,00 euro

Ejemplares sueltos:

El precio de los ejemplares atrasados es el mismo que el de los actuales.



ISSN 0210136X Dep. legal: B. 38.999 – 76

Imprime Rotocayfo-Quebecor, S.A. Ctra. de Caldes, km 3 - 08130 Santa Perpètua de Mogoda (Barcelona)

Printed in Spain - Impreso en España

TOXINAS

¿Algas destructoras del cerebro?

Las algas azulverdosas, o cianobacterias, son seguramente los organismos más extendidos, abundantes y antiguos de la Tierra. Producen una toxina de nombre BMAA vinculada a trastornos neurodegenerativos, entre ellos la enfermedad de Alzheimer y una patología de la isla de Guam similar a las enfermedades de Lou Gehring y Parkinson. Un equipo internacional ha examinado las cianobacterias que medran en aguas dulces, marinas y ligeramente salobres de todo el mundo, así como las que viven en simbiosis con plantas y líquenes. El noventa por ciento de las 41 variedades estudiadas sintetizaba BMAA; en momentos adecuados, o en ciertas fases de su desarrollo, todas podrían producirla. Llamó la atención la concentración de toxina en tapetes cianobacterianos oceánicos y del mar Báltico. Puesto que la contaminación del agua y el aumento de la temperatura global fomentan esas floraciones, que cubren miles de kilómetros cuadrados, las repercusiones para la salud podrían resultar cada vez más preocupantes.

—Charles Q. Choi



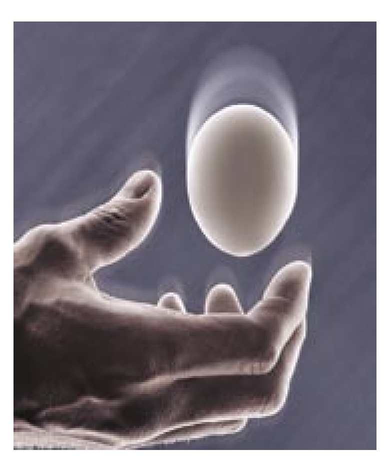
Trichodesmium thiebautil, cianobacteria del Caribe que produce neurotoxinas.

PERCEPCION

Gravedad cerebral

Los astronautas saben que se encuentran en un lugar de microgravedad. Pese a ello, cuando en la nave cae un objeto y quieren recuperarlo, alargan espontáneamente la mano, con premura excesiva, como si esperasen que el peso fuese el mismo que en tierra. Esa expectativa quizá se halle integrada de modo indeleble en el cerebro. En un ensayo realizado *ad hoc*, un equipo de investigadores italianos mostró a los voluntarios dos grupos de animaciones. En el primero, una pelota saltaba hacia arriba y volvía hacia abajo con una aceleración normal, *g*; en el otro grupo, la pelota se aceleraba hacia arriba en cada rebote. Los sujetos del experimento predecían mejor el movimiento de la bola cuando su comportamiento coincidía con el de la gravedad en tierra. Las exploraciones cerebrales acometidas mientras se realizaba el ensayo revelaron que se excitaba la corteza vestibular, que controla el sentido del equilibrio y la conciencia de la posición del cuerpo. Proponen los investigadores que en esa zona del cerebro se representa el efecto de la gravedad comparando el movimiento de la cabeza en diferentes direcciones y enviando sus cálculos a los sistemas visual y motor.

—J. R. Minkel



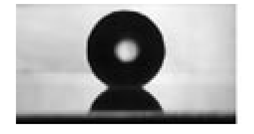
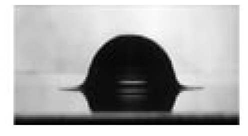
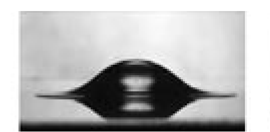
La aceleración de la gravedad podría ser una percepción indeleblemente integrada en el cerebro.

FISICA

Líquidos que no salpican

Una buena salpicadura depende sobre todo de la presión atmosférica. Cuando una gota cae sobre una superficie plana, forma al extenderse un charco ondulante que se descompone en salpicaduras. Con la pretensión de llegar a controlar el fenómeno, unos físicos de la Universidad de Chicago dejaron caer dentro de una cámara de vacío gotas de alcohol sobre una placa de vidrio, seca y lisa. Grabaron los resultados con una cámara de 47.000 fotogramas por segundo. A una presión de alrededor de un sexto de la atmosférica normal, la salpicadura desaparecía completamente; las gotas sólo adquirían forma de torta, sin ondulaciones visibles. Sospechan los investigadores que las gotas que caen, salpican porque la presión gaseosa que actúa sobre ellas desestabiliza su expansión hacia afuera. Tales observaciones podrían servir para controlar mejor las rociadas en el quemado de combustibles y en la impresión por chorro de tinta.

—Charles Q. Choi


Una gota que se dispersa y no salpica. Está sometida a una presión atmosférica de 17,2 kilopascal, un sexto de la normal.

NATIONAL ACADEMY OF SCIENCES, 2005 *(arriba)*; LESTER LEFKOWITZ/PHOTOGRAPHER'S CHOICE/GETTY *(centro)*; LEI XU *Universidad de Chicago (abajo)*

SOCIOLOGIA DE LA CIENCIA

Bizantina tolerancia de los fallos

El físico Alan Sokal, de la Universidad de Nueva York, creyó haber demostrado que algunas ramas del saber académico —de letras— se habían despeñado por un abismo de irracionalidad y absorbían cualquier texto absurdo que se les presentase mientras contuviera la jerga y los giros convenientes. Las tornas se volvieron contra la física teórica con los hermanos Bogdanoff, quienes lograron que revistas profesionales y científicos de algún prestigio dieran el visto bueno a artículos, e incluso una tesis doctoral, no mejores que el de Sokal. Ahora, unos estudiantes de doctorado del MIT han repetido la maniobra, pero en el campo de las ciencias de la computación y mediante un procedimiento extremo, aunque muy oportuno teniendo en cuenta que se trata de informáticos: colaron un artículo generado aleatoriamente por un programa de ordenador en el inminente noveno Multicongreso Mundial de Sistémica, Cibernética e Informática. Pretendían así denunciar ciertas reuniones más o menos científicas que les parecen fraudulentas. El Generador Automático de Artículos quizás haya dado más en el blanco que Sokal. Ha generado el título de esta nota. No está elegido adrede, el programa no se deja manipular. Debe de tratarse de la opinión, críptica, concisa, de las máquinas ante este fenómeno humano: una atención finita y propensa a la dejadez se siente insegura al discriminar en un entorno complejo; la imitación del estilo de una disciplina ya aporta alguna inteligencia del texto engañoso. La moneda falsa, por definición, se parece a la buena, y la buena no siempre lo es tanto.

Rooter: A Methodology for the Typical Unification of Access Points and Redundancy

Jeremy Stribling, Daniel Aguayo and Maxwell Krohn

ABSTRACT

Many physicists would agree that, had it not been for congestion control, the evaluation of web browsers might never have occurred. In fact, few hackers worldwide would disagree with the essential unification of voice-over-IP and public-private key pair. In order to solve this riddle, we confirm that SMPs can be made stochastic, cacheable, and interposable.

I. INTRODUCTION

Many scholars would agree that, had it not been for active networks, the simulation of Lamport clocks might never have occurred. The notion that end-users synchronize with the investigation of Markov models is rarely outdated. A theoretical grand challenge in theory is the important unification of virtual machines and real-time theory. To what extent can web browsers be constructed to achieve this purpose?

Certainly, the usual methods for the emulation of Smalltalk that paved the way for the investigation of rasterization do not apply in this area. In the opinions of many, despite the fact that conventional wisdom states that this grand challenge is continuously answered by the study of access points, we believe that a different solution is necessary. It should be noted that Rooter runs in $\Omega(\log \log n)$ time. Certainly, the shortcoming of this type of solution, however, is that compilers and superpages are mostly incompatible. Despite the fact that similar methodologies visualize XML, we surmount this issue without synthesizing distributed archetypes.

We question the need for digital-to-analog converters. It should be noted that we allow DHCP to harness homogeneous epistemologies without the evaluation of evolutionary programming [2], [12], [14]. Contrarily, the lookaside buffer might not be the panacea that end-users expected. However, this method is never considered confusing. Our approach turns the knowledge-base communication sledgehammer into a scalpel.

The rest of this paper is organized as follows. For starters, we motivate the need for fiber-optic cables. We place our work in context with the prior work in this area. To address this obstacle, we disprove that even though the much-tauted autonomous algorithm for the construction of digital-to-analog converters by Jones [10] is NP-complete, object-oriented languages can be made signed, decentralized, and signed. Along these same lines, to accomplish this mission, we concentrate our efforts on showing that the famous ubiquitous algorithm for the exploration of robots by Sato et al. runs in $\Omega((n + \log n))$ time [22]. In the end, we conclude.

II. ARCHITECTURE

Our research is principled. Consider the early methodology by Martin and Smith; our model is similar, but will actually overcome this grand challenge. Despite the fact that such a claim at first glance seems unexpected, it is buffetted by previous work in the field. Any significant development of secure theory will clearly require that the acclaimed real-time algorithm for the refinement of write-ahead logging by Edward Feigenbaum et al. [15] is impossible; our application is no different. This may or may not actually hold in reality. We consider an application consisting of n access points. Next, the model for our heuristic consists of four independent components: simulated annealing, active networks, flexible modalities, and the study of reinforcement learning.

We consider an algorithm consisting of n semaphores. Any unproven synthesis of introspective methodologies will clearly require that the well-known reliable algorithm for the investigation of randomized algorithms by Zheng is in Co-NP; our application is no different. The question is, will Rooter satisfy all of these assumptions? No.

Reality aside, we would like to deploy a methodology for how Rooter might behave in theory. Furthermore, consider the early architecture by Sato; our methodology is similar, but will actually achieve this goal. despite the results by Ken

El congreso WMSCI2005 aceptó como artículo original este texto, escrito en realidad por el programa SCIgen. Cuando se supo que era una broma, se retiró la aceptación.

ASTROFISICA

Erupciones breves y magnetares

El 27 de diciembre de 2004 se detectó la explosión cósmica más brillante —tal y como se la captó en la Tierra— jamás observada. Ocurrió en nuestra propia galaxia, a 50.000 años luz de distancia. Durante menos de un segundo nos llegó de ese fogonazo más energía que de la Luna llena. Vino en forma de rayos gamma. Se originó en la estrella de neutrones SGR 1806-20, del tipo "magnetar": su campo magnético multiplica por mil billones el de la Tierra (si estuviese de nosotros a la distancia de la Luna, nos sacaría las llaves del bolsillo). Se atribuye la erupción a un seísmo estelar, que reorganizó ese campo magnético. La estrella no se destruyó. En los cielos se producen numerosas erupciones de rayos gamma. Según duren más o menos de dos segundos, se las denomina de larga o corta duración. Las erupciones de larga duración van seguidas por un rescoldo en rayos X y luz visible que ha permitido estudiarlas a fondo. Ocurrieron hace miles de millones de años en galaxias jóvenes y se atribuyen a la conversión de una estrella de masa muy grande en un agujero negro. Su energía multiplica un millón de veces la de SGR 1806-20, pero su lejanía hace que aquí resulten mucho menos intensas. No se ha observado rescoldo alguno en las erupciones de corta duración; por eso mismo se las conoce mucho peor. El equipo de David Palmer, de Los Alamos, aduce que si SGR 1806-20 hubiese ocurrido en otra galaxia, no demasiado distante, se habría asemejado a una erupción de rayos gamma de corta duración. También ha calculado, basándose en intensidades y frecuencias de aparición, que sólo un cinco por ciento de esas erupciones podría ser un destello de magnetar. Según la hipótesis dominante sobre el origen de las erupciones breves, se producirían al fundirse, en una, dos estrellas de neutrones. Del resultado de Palmer se seguiría, pues, que al menos ese cinco por ciento tendría otro origen. "Y está por ver aún", comenta, "que haya alguna que emerja por fusión de estrellas de neutrones". Gracias al satélite Swift, dedicado a la observación de las erupciones, que ha empezado a funcionar este año, "la próxima que se detecte podría aclarar las cosas".

Representación imaginaria de la propagación del destello de rayos gamma de SGR 1806-20, la más poderosa jamás observada.

—*Govert Amersfoot*

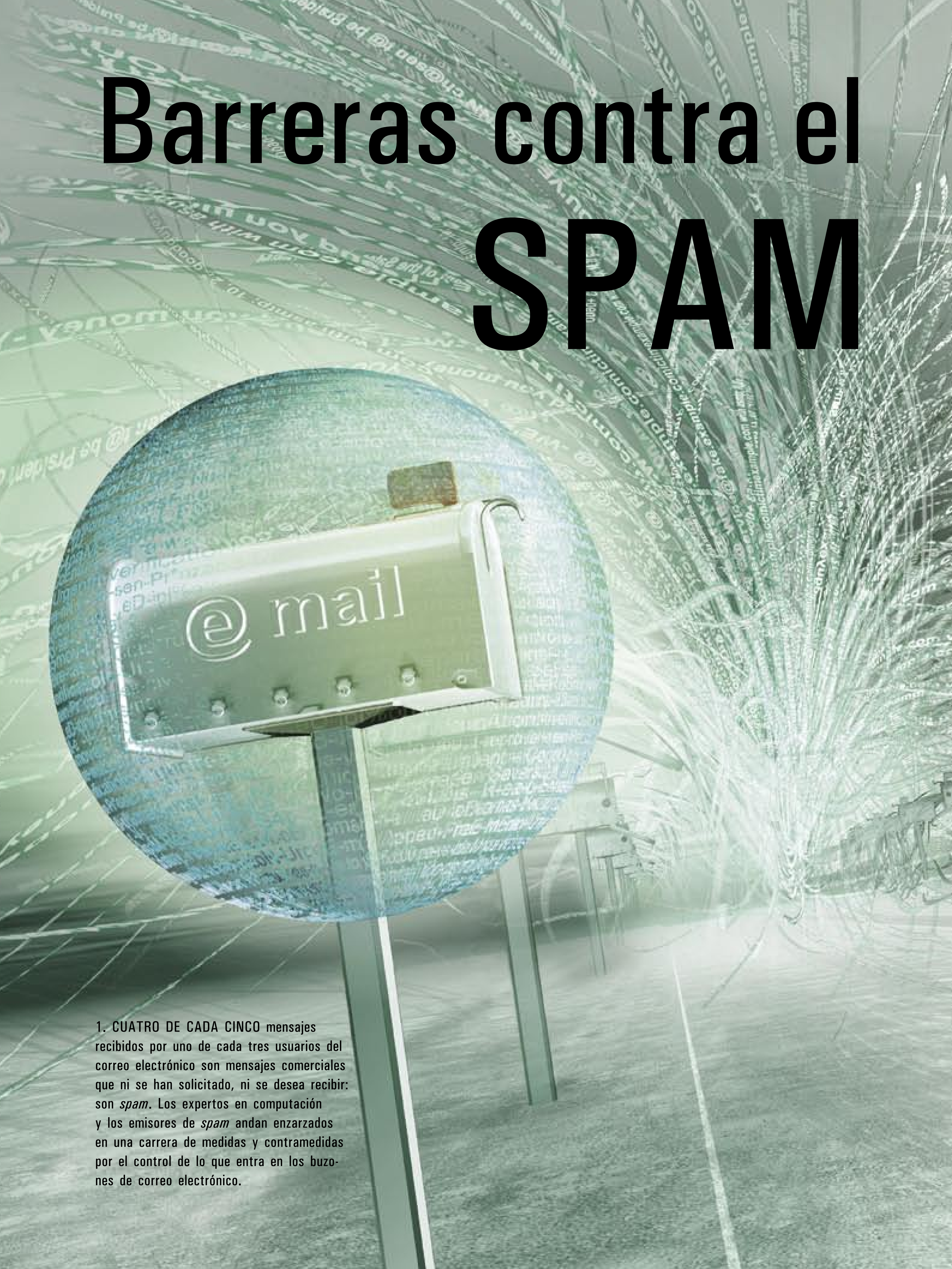
http://pdos.csail.mit.edu/scigen/ (*arriba*); NASA (*abajo*)

HERMENEUTICA CIENTIFICA

Dicho y hecho

El 9 de mayo se anunciaba que el satélite Swift había presenciado el "nacimiento de un agujero negro". Este titular debe interpretarse así: detectó una erupción de rayos gamma brevísima y, gracias a la célere reacción de sus instrumentos, un rescoldo muy débil de rayos X. Por su localización, es posible que se haya producido en una galaxia que se encuentra a más de dos mil millones de años luz de nosotros y donde apenas nacen estrellas. En tal caso, no podía tratarse del destello de un magnetar; no habría tenido energía suficiente para que, a esa distancia, lo detectase Swift. Por lo tanto, quizá se trate de la otra posibilidad en que se piensa: la fusión de dos estrellas de neutrones y su conversión en agujero negro, impresión reforzada por la inactividad de la galaxia; las estrellas gaseosas que se convierten en agujeros viven poco.

Barreras contra el SPAM



1. CUATRO DE CADA CINCO mensajes recibidos por uno de cada tres usuarios del correo electrónico son mensajes comerciales que ni se han solicitado, ni se desea recibir: son *spam*. Los expertos en computación y los emisores de *spam* andan enzarzados en una carrera de medidas y contramedidas por el control de lo que entra en los buzones de correo electrónico.

¿Qué puede hacerse para detener la oleada de mensajes indeseados en el correo electrónico?

Joshua Goodman, David Heckerman y Robert Rounthwaite

En 1978, unas 400 personas recibieron a través de Arpanet el primer correo electrónico no solicitado y quizá no deseado, el primer *spam*: una cuña publicitaria del nuevo ordenador DECSYSTEM-20 de Digital Equipment Corporation. (Aunque algunos la conviertan en un acrónimo, la palabra *spam* procede de un famoso número cómico del grupo Monty Python, emitido por la BBC en 1970, acerca de un restaurante en el que todos los platos incluían carne enlatada de la marca SPAM, una marca real, y un grupo de vikingos cantaba con entusiasmo loas a SPAM.) Actualmente circula por Internet una correspondencia basura, en forma de proposiciones comerciales no solicitadas, que puede cifrarse en miles de millones de mensajes al día, dos terceras partes del total transmitido por la Red. Un tercio de los usuarios recibe *spam* en el 80 por ciento de su correo electrónico. Ultimamente, el *spam* ha adquirido un cariz amenazante: proliferan los ataques de estafadores que expiden mensajes en los que suplantan a personas o instituciones dignas de confianza para obtener el número de la tarjeta de crédito u otros datos privados del destinatario (actividad a la que se llama *phising,* palabra de jerga derivada de *fishing*, pescar). Según un estudio de Gartner Research en 2004, estos ataques cuestan unos 1200 millones de dólares al año.

El *spam* no se limita a los mensajes de correo electrónico: en las salas de conversación por Internet (*chats*) se ocultan "robots" que pretenden ser humanos e intentan convencer a los participantes en el *chat* para que accedan a sitios pornográficos de la Red. Los usuarios de la mensajería instantánea (IM) padecen los llamados spIM, emparentados con los *spam* del correo electrónico. Para desviar las operaciones de búsqueda en Internet, se alteran las bitácoras añadiendo enlaces engañosos a ciertas hojas y así distorsionar las clasificaciones de la importancia de los sitios y enlaces de la Red.

El asfixiante efecto del *spam* parece a veces capaz de minar, cuando no destruir, las comunicaciones por Internet que hoy conocemos. La realidad no es tan sombría, sin embargo. Ya se han descubierto técnicas para interceptar mensajes *spam* y desalentar a sus autores. Pronto se unirán otras. Los métodos que examinaremos apuntan al correo basura, pero muchos de ellos servirían también para otras clases de *spam*. Ninguno de ellos será un talismán, pero su combinación, utilizada por suficiente número de usuarios, podría lograr maravillas. En absoluto es una quimera esperar que nuestros buzones de correo electrónico se vean un día libres —casi del todo— de esta enojosa plaga.

Mensajes insidiosos

La proliferación del correo electrónico indeseado es resultado directo de las tendencias del mercado que lo favorecen: la distribución de mensajes basura es baratísima. Pero no gratuita: el envío de un mensaje cuesta del orden de una centésima de céntimo. Aunque el porcentaje de respuestas fuese bajísimo (uno entre 100.000), un emisor de *spam* todavía sacaría algún beneficio con que ingresase 11 euros por artículo que vendiese. De ahí que, aunque sean muy pocos los usuarios de e-mail que compren algo anunciado por *spam*, todos padezcamos las consecuencias.

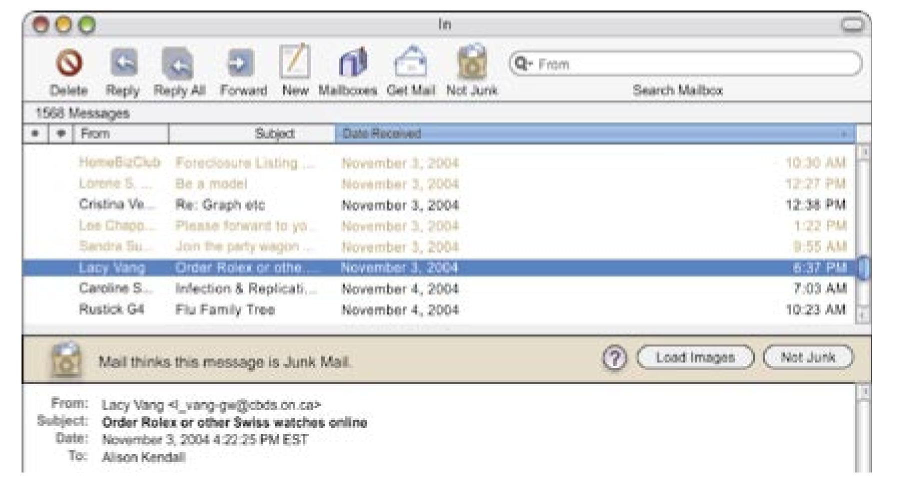
Uno de los aspectos más irritantes del *spam* es su continua adaptación a los nuevos intentos de acabar con él. Cada vez que los informáticos se oponen por algún procedimiento al *spam*, quienes lo difunden encuentran el modo de esquivar la defensa. Esta carrera ha conducido a una continua evolución en uno y otro lado, a un refinamiento cada vez mayor de las medidas y contramedidas.

Otro problema fundamental deriva de la extrema dificultad de definir el *spam*, tanto técnica como legalmente. Legalmente se lo suele describir de esta forma: "mensaje de correo electrónico comercial no solicitado procedente de alguien con el que no existe relación profesional anterior". Esta definición, si embargo, es demasiado amplia. Una reciente propuesta recibida por e-mail nos invitaba a convertir en película de cine un cuento breve que habíamos publicado en Internet. La comunicación cumplía los requisitos de la definición: no había sido solicitada, era comercial y su expedidor era desconocido, pero casi nadie la consideraría *spam*. Otra definición podría tener en cuenta que el *spam* se distribuye masivamente. Pero hace poco hemos solicitado contribuciones a un congreso sobre sistemas de correo electrónico y métodos encaminados a impedir el *spam*, para lo cual nos dirigimos a cincuenta personas que habían publicado sobre este asunto. Nunca las habíamos tratado antes, pero ninguna de ellas se quejó. Quizás el *spam* se caracterice sobre todo por su carácter indiscriminado y por no ser deseado. Igual que sucede con la pornografía, cuesta mucho definir con precisión el *spam*, pero ciertamente lo reconocemos cuando inunda nuestros buzones.

Morfología de mensajes

Trabajamos sobre los mensajes basura desde 1997, cuando uno de nosotros (Heckerman) sugirió utilizar como eficaz línea de ataque métodos de aprendizaje automático. Desde entonces, a la par que numerosos colegas

JEAN-FRANCOIS PODEVIN



2. **LOS PROGRAMAS DE CONTROL DE ACCESO identifican el correo basura mediante la localización y la asignación de pesos a características que los usuarios han considerado indeseables en casos anteriores.**

ALISON KENDALL

en el campo de la informática, hemos investigado y desarrollado diversos procedimientos para frenar el *spam*. En ellos se combinan soluciones técnicas y legales, junto con iniciativas que habrían de abarcar todo el campo de la informática.

Entre las técnicas más antiguas figuran las de comparación de huellas: consisten en obtener ejemplos de mensajes basura y elaborar una "huella digital" de los mismos mediante un programa informático. La "huella" sería un número deducido del contenido del mensaje, de tal modo que todos los mensajes idénticos o similares se identificaran por el mismo número. Como ejemplo sencillo, podría sumarse al número de veces que aparezca la letra A en un mensaje el de letras B multiplicado por 10, el de letras C por 100, etcétera. Cuando llega un nuevo mensaje, los programas *antispam* calculan su huella y la comparan a la de otros correos *spam* conocidos. Si las huellas coinciden, el programa elimina o archiva el mensaje.

Por desgracia, los autores de mensajes basura traspasaron fácilmente estas sencillas protecciones: les bastó con incluir en los mensajes unos caracteres aleatorios. A ello se respondió elaborando huellas más depuradas que excluían las secuencias de caracteres obviamente aleatorios, pero los atacantes anularon tales defensas dando una apariencia más regular al contenido aleatorio, por ejemplo la de falsos partes meteorológicos. Resulta difícil, en estos últimos tiempos, construir sistemas de identificación por huella suficientemente inmunes a la inserción de secuencias aleatorias en correos basura.

Filtros inteligentes

En vez de seguir la senda de las huellas digitales, nuestro grupo prefirió aprovechar la capacidad de aprendizaje automático de los ordenadores. Los programas informáticos especializados pueden aprender a distinguir los mensajes basura de los que sean válidos, y no se dejan confundir tan fácilmente por la adición de letras o palabras aleatorias.

Ante todo, ensayamos los métodos más sencillos y comunes de aprendizaje automático. El algoritmo Naive Bayes empieza por asignar probabilidades a cada palabra de un mensaje. Por ejemplo, "clic", "aquí" y "suscribirse" podrían tener cada una de ellas una probabilidad 0,9 de aparecer en un mensaje *spam*, y 0,2 si el mensaje fuera un e-mail legítimo. Multiplicando las probabilidades de todas las palabras contenidas en un mensaje y aplicando el principio estadístico denominado regla de Bayes, se obtiene una estimación de la probabilidad de que dicho mensaje sea basura.

La estrategia de Naive Bayes funciona muy bien para determinar la apariencia de un mensaje auténtico; como todos estos métodos de aprendizaje, resiste los intentos sencillos de confundir. Conocemos bien, sin embargo, sus limitaciones. Da por supuesto que las palabras de los e-mail son independientes y no están relacionadas, lo cual en muchas ocasiones es falso (por ejemplo, "clic" y "aquí" a menudo van unidas). Así se distorsiona el resultado.

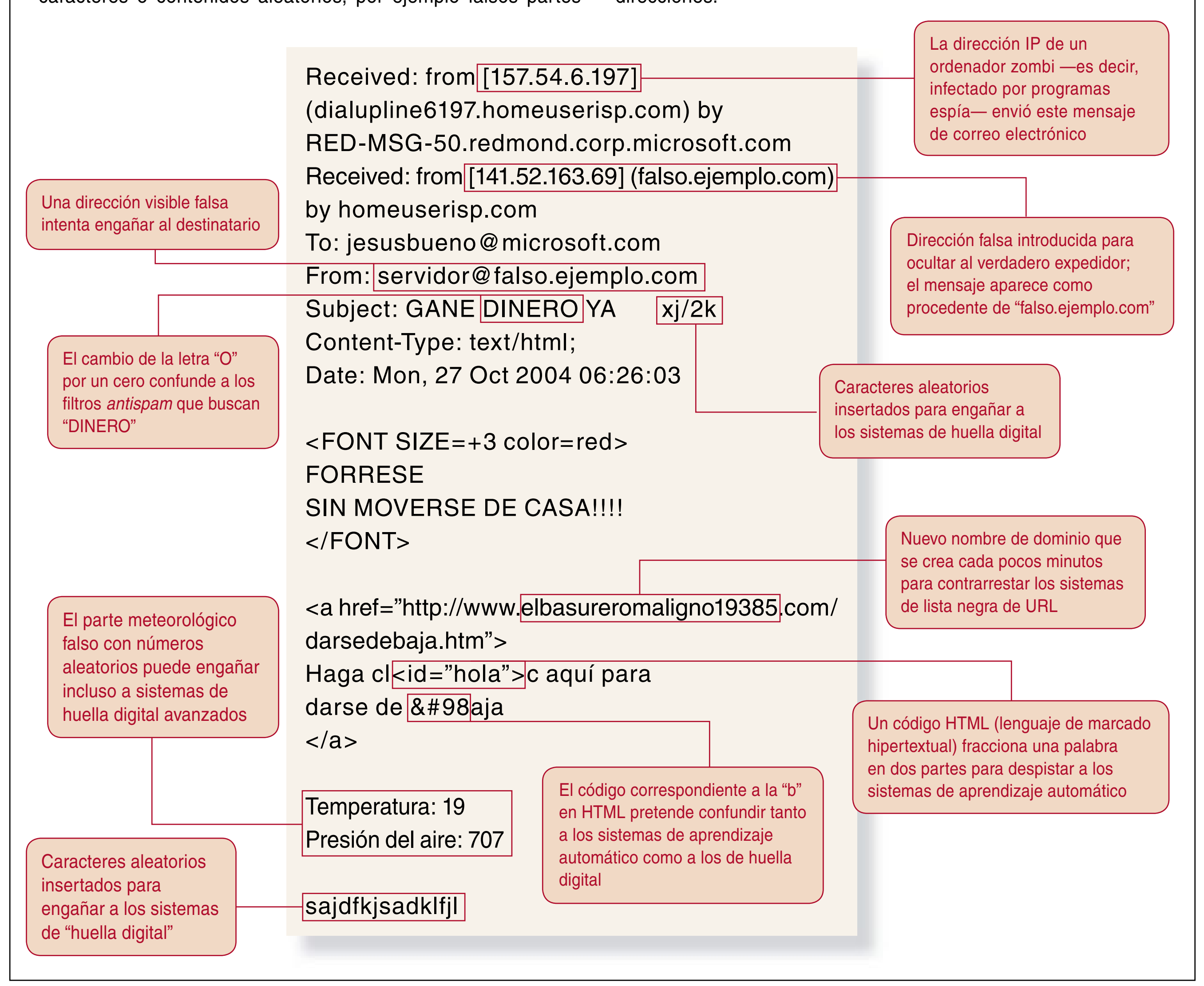
Ante tales dificultades, nos hemos centrado en los modelos lineales discriminadores, que optimizan sus decisiones posteriores mediante la asignación de pesos a los distintos elementos de juicio: a las palabras y a otras particularidades de los mensajes, como el que se haya enviado a numerosos destinatarios. Estos modelos pueden "aprender" las relaciones que ligan las palabras; por ejemplo, "sabiendo" que no debe darse demasiado peso a palabras que tienden a agruparse, como las ya citadas "clic", "aquí" y "suscribirse". Para entenderlo mejor, imaginemos que un modelo Naive Bayes haya visto estas tres palabras, a menudo asociadas a un correo *spam*. Tal vez decida que basta para calificar como basura cualquier mensaje que las contenga, lo que llevaría a destruir correo válido. Por el contrario, un modelo enseñado a discriminar debería saber que estas palabras suelen venir juntas y, por lo tanto, les asignaría unos pesos menores, más razonables. Hasta podría aprender que una palabra tal como "aquí", acaso más frecuente en los correos *spam*, no debe recibir peso alguno, pues realmente no ayu-

Resumen/Proteger el buzón de entrada

- La oleada creciente de mensajes electrónicos no deseados —*spam*— amenaza las comunicaciones en Internet. Los programadores informáticos sostienen una guerra continua contra los expedidores de correo basura, que no dejan de superar las barreras que se les interponen.
- Un conjunto de medidas *antispam*, ya conocidas y de nuevo cuño, que comprenda filtros informáticos inteligentes, sistemas que certifican la legitimidad de los expedidores de correo electrónico y rigurosas prohibiciones legales, podrían detener el torrente de mensajes indeseados si se aplicaran de manera generalizada.

TRETAS DEL *SPAM*

Se utilizan diversos métodos para eludir los filtros antibasura. Una de las medidas más elementales contra el *spam* es la de contrastar “huellas”: los ordenadores eliminan los mensajes nuevos cuyas características, su “huella”, coincida con la de mensajes *spam* conocidos que han analizado previamente. Pero los creadores de *spam* aprendieron pronto a superar el simple cotejo de huellas mediante la adición de caracteres o contenidos aleatorios, por ejemplo falsos partes meteorológicos, que modifican la huella. Más tarde empezaron a enmascarar las palabras que suelen vincularse al *spam*, escribiendo, digamos, “DINERO” como “DINER0” (con el dígito “0” en vez de la letra “O”). También hay filtros que buscan enlaces insertos en los mensajes a páginas Web y servidores de notoria relación con el *spam*. Sin embargo, los creadores de basura saben generar continuamente nuevas direcciones.



ALISON KENDALL

da a discernir lo bueno de lo malo. Además, los métodos discriminadores quizá descubrirían que ciertas palabras se cancelan mutuamente. Por ejemplo, la palabra “húmedo” es más frecuente en *spam*, pero si va unida a “tiempo” es probable que el mensaje sea legítimo.

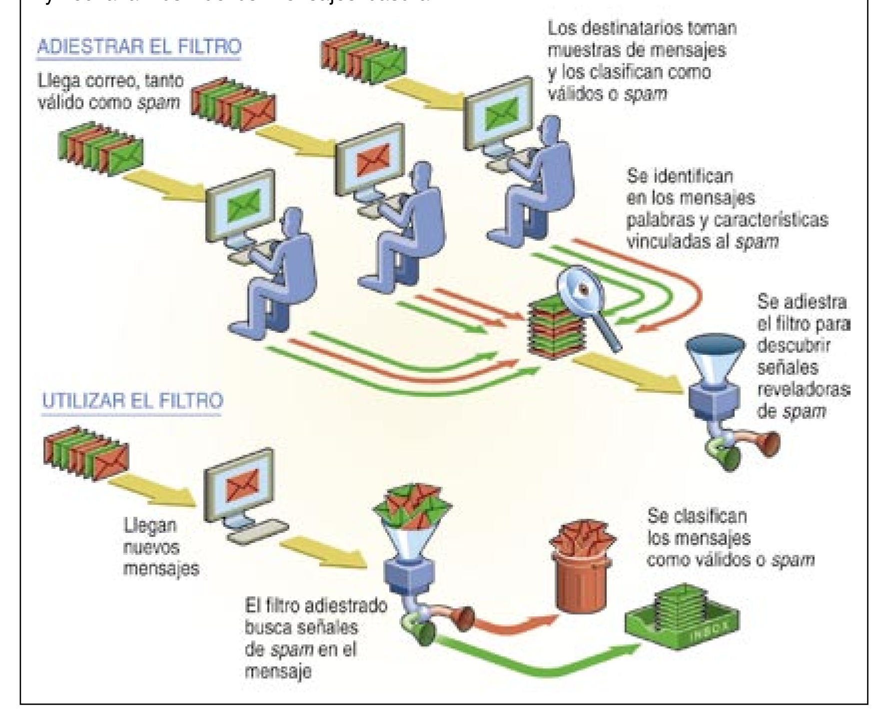
Una ventaja de los sistemas Naive Bayes es que aprenden con facilidad. La determinación de pesos para los métodos discriminadores es mucho más ardua: exige que los programadores ensayen muchos conjuntos de valores de pesos para las palabras y las demás características, buscando una combinación que optimice la discriminación entre *spam* y mensajes legítimos. Afortunadamente, los avances han sido notables en este terreno. Algoritmos como el de optimización secuencial mínima, de John C. Platt, de Microsoft, y el de escalamiento iterativo secuencial condicional generalizado (SCGIS, por sus siglas en inglés), inventado por uno de nosotros (Goodman), son decenas o cientos de veces más rápidos que los de técnicas anteriores. Cuando se manejan grandes cantidades de datos para el aprendizaje *antispam*, más de un millón de mensajes y cientos de miles de pesos asignados, la rapidez del algoritmo es de suma importancia.

Esconder la basura

De sobra sabíamos que nuestros sistemas de aprendizaje automático, centrados en las palabras del mensaje,

USUARIOS DEL CORREO ELECTRONICO ADIESTRAN FILTROS CONTRA EL *SPAM*

Los primeros filtros de *spam* producidos por Microsoft se basaban en datos recogidos sólo de veinte usuarios del correo electrónico. A medida que los creadores de *spam* aguzaron su ingenio se necesitó una fuente de datos mejor. Cien mil voluntarios del servidor de correo Hotmail clasifican como legítimos o *spam* mensajes seleccionados aleatoriamente entre los que reciben. Con ello el sistema aprende a señalar nuevos objetivos del filtrado. Cuando los creadores de *spam* encuentra un camino para eludir el filtro, el sistema tarda muy poco en reconocer y rechazar los nuevos mensajes basura.



serían vulnerables a correos basura que enmascarasen los términos empleados. Por ejemplo, los redactores astutos de *spam* en inglés aprendieron a utilizar palabras como "M0NEY" (que sustituye la letra "O" por un cero) o trucos del HTML (lenguaje de marcado hipertextual con que se suelen escribir los documentos de la Red), como la división de una palabra en trozos ("cl" e "ick", en lugar de "click"). Al no figurar ya en el mensaje los términos acusadores ("money", "click"), es posible engañar al sistema de filtrado. Pero es bueno saber que los sistemas también pueden aprender estas artimañas y adaptarse a ellas.

Por desgracia, nos equivocamos al creer que poca gente respondería a un mensaje que claramente se valiera de medios arteros para superar el filtrado antibasura. La triste realidad es que los compradores de productos irregulares o abiertamente ilegales no esperan que los vendedores utilicen técnicas publicitarias correctas. Hemos tenido, pues, que modificar nuestros sistemas de aprendizaje empleando los modelos llamados "n-gram". Consisten en utilizar fragmentos de palabras para detectar palabras clave que suelen asociarse al *spam*. Si, por ejemplo, un mensaje de correo electrónico contuviera la frase "n@ked l@dies" ("señoras desnudas"), de la misma se extraerían los n-gram "<space>n@k", "n@ke", "@ked", etcétera. Dado que estos fragmentos aparecen en mensajes *spam* confirmados, su presencia proporciona valiosas pistas para el filtrado.

Las técnicas "n-gram" nos han ayudado a mejorar la utilidad de nuestros filtros cuando se aplican a idiomas extranjeros. El chino y el japonés, por ejemplo, no separan las palabras por espacios, lo que dificulta mucho encontrar divisiones de palabras. En tales lenguas, los sistemas equipados con n-gram examinan toda palabra y fragmento de palabra posible.

Basura en imágenes

A veces el *spam* se esconde en una imagen, cuyo contenido no pueden analizar los sistemas de aprendizaje automático, aunque todavía dispongan de otras pistas, como los enlaces presentes en el mensaje o la información sobre la reputación del expedidor. Un campo abierto a las investigaciones futuras es el filtrado por reconocimiento óptico de caracteres: las mismas técnicas de ese tipo empleadas para escanear un documento podrían localizar los textos incorporados a imágenes y llevarlos a un filtro de aprendizaje automático.

Uno de los aspectos que pueden resultar más ofensivos del *spam* es la aparición de imágenes pornográficas en el propio buzón electrónico. La visión computerizada ha hecho grandes progresos en la detección automática de ese tipo de imágenes. La labor a realizar sorprende por su amplitud: desde impedir el acceso infantil a sitios Web que contengan material sexual hasta evitar que los productores de pornografía abusen de los servidores gratuitos de la Red. Con todo, el reconocimiento de imagen es todavía lento y es preciso mejorar la seguridad de la identificación. Imágenes inocuas, sobre todo las que muestran grandes extensiones de piel, podrían dar falsos positivos.

Nuestro grupo investiga asimismo cómo reconocer los mensajes de correo basura por análisis del URL (localizador universal de recursos, el código que remite a otras páginas Web). La mayoría de los difusores de *spam* persigue en primer lugar que los usuarios visiten su sitio en la Red, aunque algunos prefieran el contacto a través del número telefónico. Por tanto el URL es especialmente valioso para el filtrado.

Hay muchas maneras de aprovechar la información contenida en el URL. Ciertos proveedores de programas *antispam* han empezado ya a bloquear mensajes que contienen enlaces a páginas Web a todas luces relacionadas con el correo basura. Los enlaces a dominios no conocidos previamente despiertan sospechas: los difusores de *spam* crean con gran rapidez dominios, mientras que la mayoría de los dominios legítimos perduran largo tiempo. Por

SAMUEL VELASCO

otra parte, el URL puede también ser indicador de correo legítimo: si un mensaje sólo contiene punteros hacia páginas conocidas que no tengan relación con *spam*, o carece en absoluto de URL, resulta menos probable que sea basura.

Pasar la comprobación

Las técnicas de filtrado funcionan bastante bien, pero sabemos que los creadores de *spam* siempre intentarán atacarlas. En vez de pretender ganar esta competición inacabable, creemos que a largo plazo lo más práctico sería cambiar las reglas del juego. Por eso estamos examinando sistemas de comprobación, orientados a exigir del productor del *spam* más de lo que pueda aguantar.

Para enviar el primero de todos los mensajes *spam* hubo que teclear 400 direcciones de correo electrónico. Hoy día casi todo el correo basura se envía de modo automático. Si un expedidor puede demostrar que es humano, probablemente el mensaje no será *spam*. Esta idea se aplicó en uno de los sistemas de comprobación más antiguos, ideado por Moni Naor, del Instituto Weizmann de Ciencias, en Israel. Proponía utilizar las que han venido a llamarse comprobaciones humanas interactivas (HIP, por sus siglas en inglés), CAPTCHA —acrónimo inglés de "prueba pública Turing enteramente automatizada para distinguir personas de ordenadores"— o pruebas Turing invertidas. La HIP es un problema o jeroglífico concebido para que las personas lo resuelvan con facilidad, pero de máxima dificultad para los ordenadores. Las personas, por ejemplo, aventajan mucho a las máquinas en el reconocimiento de caracteres alfabéticos aleatorios parcialmente ocultos o distorsionados dentro de una imagen.

La comprobación HIP forma parte de un sistema de problemas y respuestas que ha de comprobar que el expedidor es humano. Antes de entregar el mensaje, el sistema revisa una "lista blanca" de expedidores que el destinatario considere dignos de confianza. Si el expedidor figura en la lista, el mensaje se depositará en el buzón de destino. En caso contrario, se emitirá un mensaje al expedidor original en el que se le pedirá que resuelva una HIP. Si lo consigue, la respuesta se enviará al destinatario, cuyo programa de correo electrónico transferirá inmediatamente el mensaje a su buzón de entrada.

Este tipo de interactividad puede, sin embargo, resultar enojosa para los usuarios. A poca gente le gusta tener que resolver una HIP para enviar correos electrónicos; algunos incluso se niegan a hacerlo. Naor y Cynthia Dwork han sugerido un procedimiento

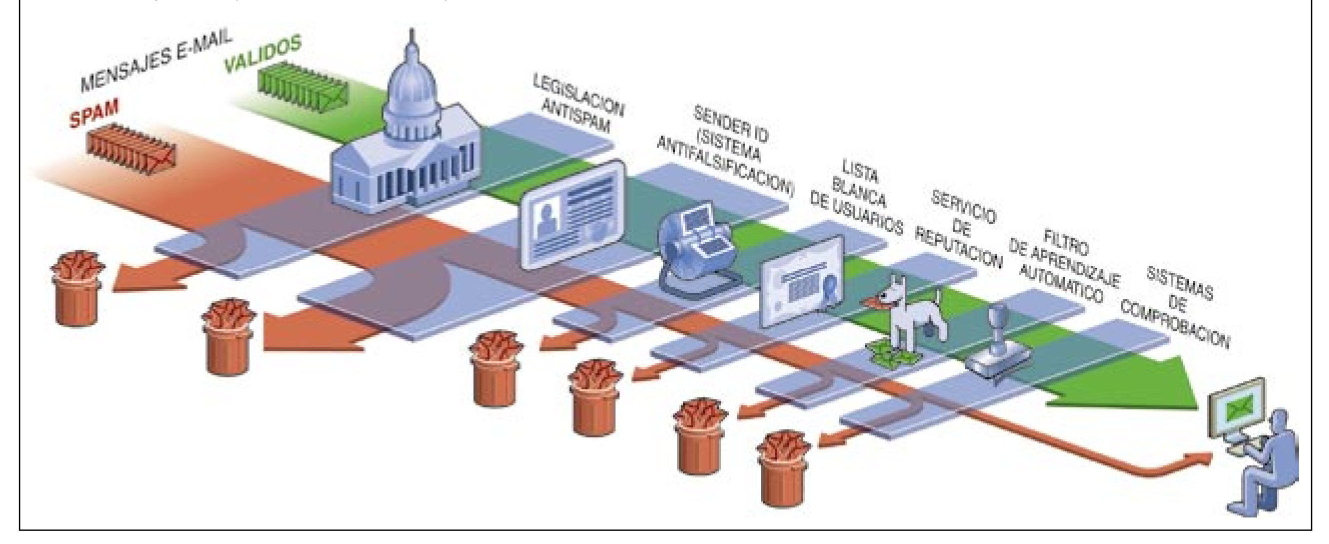
MEDIDAS MULTIPLES CONTRA EL CORREO BASURA

Interceptar el *spam* requiere una defensa multiforme que combine leyes disuasorias, técnicas para descubrir direcciones falsas de expedidores, el filtrado inteligente de los mensajes y sistemas de comprobación que verifiquen que el mensaje procede de una persona física o encarezcan excesivamente la distribución de *spam*.

La primera línea de defensa es la legislación. En Estados Unidos, la ley CAN-SPAM prohíbe ciertas prácticas particularmente detestables, aunque hasta ahora no parece que haya arredrado a la mayoría de los creadores de *spam*. En la Unión Europea están en vigor la "Directiva sobre el comercio electrónico" y leyes nacionales adaptadas a esa norma, como, en España, la Ley de Servicios de la Sociedad de la Información.

Puesto que cerca de la mitad de mensajes *spam* utilizan direcciones de expedidor falsas, el nuevo marco de identificación del expedidor (*Sender ID Framework*) añade al servidor de nombres de dominio (DNS) información suplementaria que enumera las direcciones de protocolo Internet (IP) de los ordenadores autorizados a expedir correo desde ese dominio.

El sistema de correo electrónico de un PC verifica una "lista blanca" de expedidores dignos de confianza, seleccionados previamente por el destinatario. Los incluidos en la lista evitan el filtrado posterior; los no incluidos sufrirán el examen de un servicio de reputación destinado a los expedidores que hayan acordado respetar normas estrictas para impedir la generación de *spam*. Si el expedidor no figura en ninguna de las dos listas, se dirige el mensaje a un filtro *antispam* basado en el aprendizaje automático. Ante una mínima sospecha, se exigirá al expedidor que aporte una prueba: resolver un sencillo acertijo para cerciorarse de que es una persona física, responder un problema más complicado que consuma un costoso tiempo de ordenador o depositar una pequeña cantidad reembolsable.



SAMUEL VELASCO

sistemas de reconocimiento de imagen ANTIPORNOGRAFICOS





3. LOS SISTEMAS ANTIPORNOGRAFICOS automatizados que eliminan las imágenes sexualmente explícitas a veces malinterpretan ilustraciones inocuas, como las aquí presentadas.

alternativo, en este caso automatizado, que recurre a acertijos informáticos. La entrega correcta de un mensaje exigiría que el sistema de correo del expedidor resolviese un problema informático planteado por el sistema del destinatario. Se demostraría así que el expedidor ha invertido más tiempo de ordenador en ese mensaje que el que podría permitirse un productor de *spam* en grandes cantidades. Los acertijos informáticos, como los rompecabezas, son difíciles de resolver aunque fáciles de comprobar. En promedio, se tardarían muchos segundos o incluso minutos en solucionarlos, pero sólo milisegundos en validarlos. Para resolver estos problemas con presteza los creadores de *spam* tendrían que comprar muchos ordenadores, con una repercusión prohibitiva en sus costes.

En otro tipo de sistemas de comprobación se utiliza dinero real. Los expedidores incorporan al mensaje una especie de cheque electrónico por una cantidad irrisoria, cinco céntimos, por ejemplo, imprescindible para atravesar los filtros antibasura. Cuando el mensaje es válido su destinatario, no hace caso del cheque; mas si resulta ser *spam*, un mecanismo de denuncia normalizado le permitirá cobrar su importe o donarlo a fines benéficos. Al mismo tiempo, un mecanismo informático limita la cantidad enviada en la totalidad de mensajes que genera un expedidor, a fin de garantizar que no envíe más correos de los que admita su presupuesto. El sistema será gratuito para los expedidores legítimos; a los emisores de basura, en cambio, se les cargará un coste, un céntimo quizá, unas cien veces el precio actual de enviar el mensaje, que resultará excesivo para su economía. Los usuarios individuales dispondrían además de un pequeño depósito virtual constituido por su proveedor de Internet o al adquirir el programa de correo electrónico, de manera que para la mayoría de ellos no habría coste alguno.

Pese a la sencillez de la idea, esta clase de sistemas monetarios son de difícil realización. Los dispositivos electrónicos imponen gastos de mantenimiento; las transacciones no resultarían gratuitas. Y quedan sin contestar muchas preguntas sobre la infraestructura bancaria de los micropagos: ¿de dónde saldrá el dinero para costearla? ¿Cómo se mantendrá en servicio y a quién beneficiará? ¿Quién percibirá los pagos y cómo evitará los fraudes el sistema? Ninguno de estos problemas es insoluble, pero será arduo explotar una estructura de este género.

Un ataque integral

Nuestra estrategia preferida para frenar el *spam* combina las técnicas de filtrado del correo con un surtido de comprobaciones: las HIP, los acertijos informáticos y los micropagos. Según este enfoque, si el expedidor de un mensaje no figura en la lista blanca del destinatario, el mensaje se deriva a un filtro antibasura basado en el aprendizaje automático. El filtro es intolerante: ante la menor sospecha, somete a prueba al emisor. La mayoría de los mensajes cruzados entre personas, no obstante, se libra del examen; así se reduce radicalmente el número de comprobaciones. El expedidor original puede elegir entre resolver una HIP o un problema informático, o bien efectuar un micropago reembolsable. Si su ordenador incorpora los programas más modernos, se resolverá automáticamente el acertijo y el expedidor ni siquiera se enterará de la prueba. De no ser así, tendrá que superar una HIP o hacer un micropago.

Empresas e instituciones, sea cual fuere su magnitud, no pueden ir más allá contra el *spam*. Para una solución completa se necesitaría la cooperación de los fabricantes de ordenadores y soportes informáticos, así como de los gobiernos.

Cerca de dos tercios de los mensajes de correo electrónico falsean la dirección del expedidor. Los protocolos hoy utilizados se basan en la confianza: el expedidor dice sencillamente quién es y el destinatario le cree. Daba buen resultado en los primeros tiempos de Internet, antes de que proliferara el correo basura y se utilizase el correo electrónico para transacciones comerciales.

Cuesta mucho trabajo cambiar las normas de Internet, y en especial los protocolos del correo electrónico. Sin embargo, al fin tenemos un nuevo estándar, el "marco de identificación del expedidor" (*Sender ID Framework*), que aborda el problema de la falsificación. Consiste en añadir al servidor de nombres de dominio (DNS) información suplementaria que enumera las direcciones del protocolo Internet (IP) autorizadas a expedir correo desde un dominio especificado (parte de la red). Las direcciones IP son numéricas, como direcciones postales para cada ordenador individuales, del tipo "1.2.3.4". La nueva lista de entradas DNS para un dominio dado —digamos "ejemplo.com"— determina las direcciones IP que tienen permitido enviar correo desde ese dominio. Si alguien fingiera pertenecer a "ejemplo.com" para expedir *spam*, su dirección IP no coincidiría con ninguna de las

JOSHUA GOODMAN (*izquierda*); DAN GURAVICH *Corbis* (*derecha*)

entradas del expedidor ID de “ejemplo.com”; el programa de correo electrónico detectaría la falsificación.

Aunque conocer la identidad del expedidor sea esencial para evitar fraudes, el “lanzamiento del anzuelo” (*phishing)* no resolverá el problema del correo basura. Nadie impedirá a los generadores de *spam* fabricarse nuevas identidades todos los días o cada pocos minutos. Por eso tendrán tanta importancia los servicios de “reputación”, que permiten a los expedidores certificar su legitimidad.

En una modalidad, el programa *IronPort Bonded Sender*, el expedidor deposita dinero en concepto de fianza. Si la suma de los mensajes del expedidor denunciados rebasa un determinado umbral, el dinero de la fianza se aplicará a un fin benéfico. Los filtros *antispam* podrán revisar la lista de expedidores garantizados y dejarán pasar el correo de cualquiera de ellos, aunque despierte algunas sospechas. Estos programas también podrían actuar incluso para los que envíen pocos mensajes. Un proveedor de acceso a Internet, como MSN o AOL, podría acogerse a un servicio de reputación para obtener acceso a su programa de certificación; el proveedor vigilaría entonces el volumen de correo y los mensajes denunciados de cada uno de sus usuarios para garantizar que ninguno de ellos sea productor de correo basura.

Si la mayoría de los emisores legítimos adoptaran esta clase de sistemas (y hay muy pocas razones para no hacerlo), podrían aplicarse filtros mucho más incisivos contra el correo restante y se detendría la inmensa mayoría de los mensajes basura. Se podrían combinar los sistemas de reputación con los de pregunta y respuesta: quienes no puedan incorporarse a ellos dispondrían así de otra opción para expedir su correo.

La legislación es otra manera de detener el correo basura. En enero de 2004 entró en vigor en Estados Unidos la ley CAN-SPAM. En sí misma no ilegalizaba el correo basura; se limitaba a prohibir conductas especialmente escandalosas, como falsear el remite (el renglón From: ... en la cabecera de los mensajes). Por desgracia, esta disposición legal ha tenido muy poco efecto hasta la fecha: durante su vigencia, la proporción de correos *spam* con remites falsos ha crecido del 41 al 67 por ciento. Los países europeos, por el contrario, han aprobado leyes mucho más estrictas, que impiden enviar correos electrónicos comerciales sin el permiso expreso de su destinatario (es el caso de España, con la Ley de Servicios de la Sociedad de la Información, que entró en vigor en 2002). Esas leyes han demostrado cierta eficacia, por lo menos en detener la emisión de correos basura por las grandes empresas.

Ciertamente, no puede esperarse que ninguna ley de un solo país acabe con el correo basura. Cerca de la mitad del *spam* total procede de los EE.UU. Sólo uno de cada tres productos vendidos gracia al *spam* (seguros o refinanciación de hipotecas) requiere que la firma resida en el país del cliente. Los demás, se trate de pornografía, “hierbas vigorizantes” o de timos, se domicilian en cualquier parte o son ya ilegales.

Un futuro sin *spam*

Las firmas informáticas, los creadores de programas de dominio público y las instituciones académicas siguen estudiando cómo desterrar el correo basura. Recientemente hemos ayudado a convocar el primer congreso formal sobre el tema —el Congreso sobre el Correo electrónico y el *Antispam*— a la que acudieron investigadores de todo el mundo. Unos ingenieros de IBM presentaron la aplicación de técnicas bioinformáticas, originalmente diseñadas para encontrar estructuras genéticas, al reconocimiento de configuraciones de *spam*. Varios investigadores de AOL demostraron que múltiples sistemas de huella digital con diferentes vocabularios se defenderían mejor contra las tretas de los productores de basura. Un equipo de la Universidad de California en Davis describió la manera de atacar eficazmente los filtros *antispam* de aprendizaje automático añadiendo unas pocas palabras comunes y el modo de contrarrestar ese ataque mediante el adiestramiento de los filtros.

Nos quedan pocas dudas de que la combinación de las técnicas actuales con las de la próxima generación acabará por interceptar la mayoría del correo basura. Siempre habrá, desde luego, algún intruso que esté dispuesto a pagar el precio de entrar en nuestros buzones, pero la inundación actual se convertirá en un goteo.

Los autores

Joshua Goodman, David Heckerman y Robert Rounthwaite llevan muchos años trabajando en procedimientos que detengan el correo basura. Heckerman y Rounthwaite, junto a otros, crearon en 1997 el primer filtro *antispam* por aprendizaje automático. Heckerman dirige el grupo de aprendizaje automático y estadística aplicada (MLAS) de Microsoft Research. Goodman y Rounthwaite participaron en la organización del equipo de Microsoft que suministra las técnicas aplicadas en los sistemas de correo Exchange, Outlook, MSN y Hotmail. Rounthwaite es arquitecto jefe del grupo, y Goodman investiga dentro del MLAS el *spam* y el correo electrónico.

Bibliografía complementaria

A BAYESIAN APPROACH TO FILTERING JUNK E-MAIL. M. Sahami, S. Dumais, D. Heckerman y E. Horvitz. AAAI Technical Report WS-98-05, Madison, Wisconsin, 1998. http://citeseer.ist.psu.edu/sahami98bayesian.html

PATTERN CLASSIFICATION. Segunda edición. Richard O. Duda, Peter E. Hart y David G. Stork. John Wiley & Sons, 2000.

LEARNING TO FILTER UNSOLICITED COMMERCIAL E-MAIL. Ion Androutsopoulos, Georgios Paliouras y Eirinaios Michelakis. Technical Report 2004/2, NCSR Demokritos. http://iit.demokritos.gr/~paliourg/papers/TR2004_updated.pdf

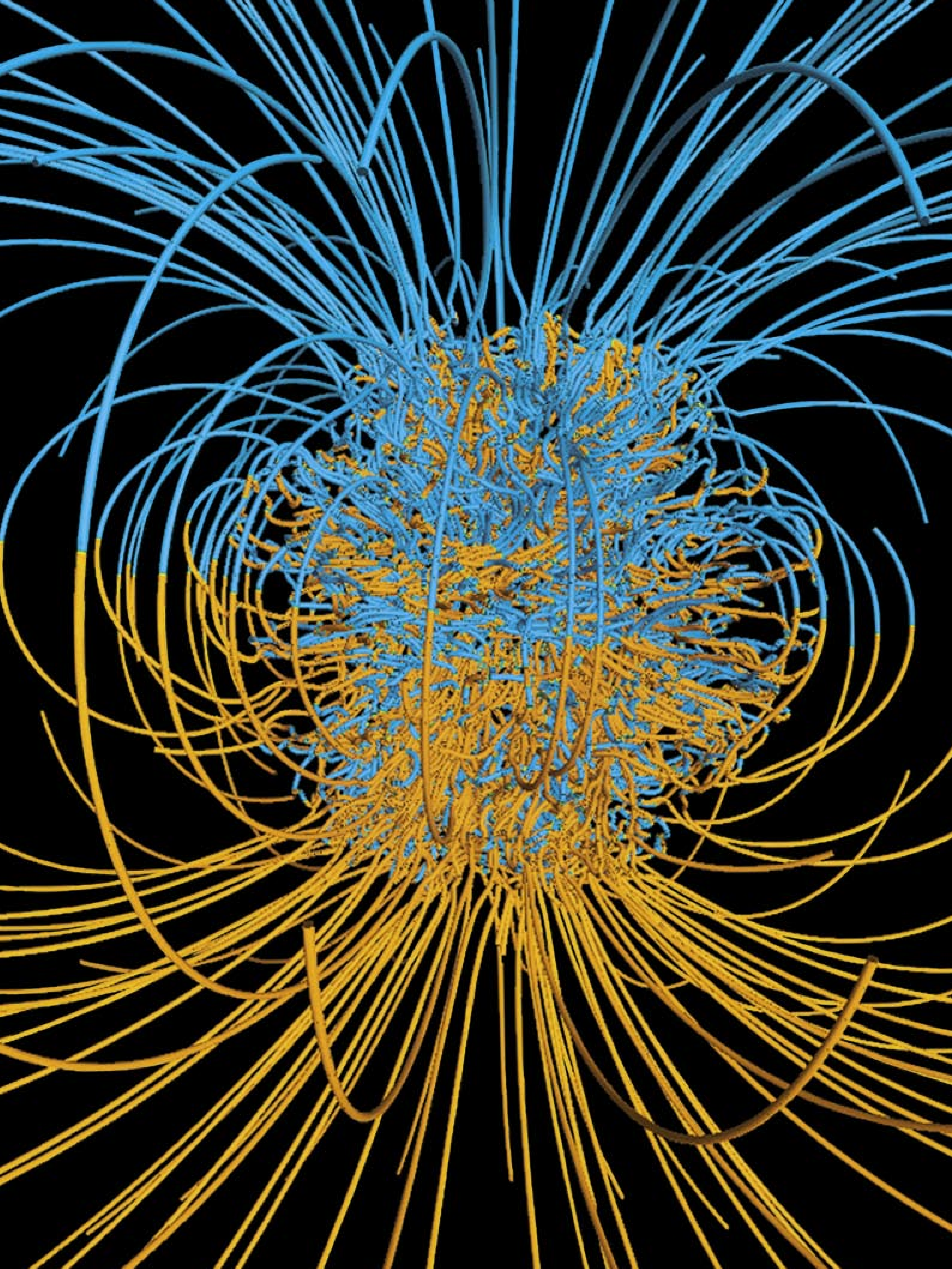
SPAM KINGS: THE REAL STORY BEHIND THE HIGH-ROLLING HUCKSTERS PUSHING PORN, PILLS, AND %*@)# ENLARGEMENTS. Brian McWilliams. O'Reilly, 2004.

Congreso sobre Email y Anti-*Spam*: www.ceas.cc

A PLAN FOR *SPAM*. Paul Graham. www.paulgraham.com/*spam*.html

SPAM: TECHNOLOGIES AND POLICIES. Joshua Goodman. www.research.microsoft.com/~Joshuago/*spam*tech.pdf

Sugerencias a los usuarios para evitar el *spam*: www.microsoft.com/athome/security/*spam*/fight*spam*.mspx



La geodinamo

Hace mucho que se quiere saber por qué se invierte la polaridad del campo magnético de la Tierra. Estudios recientes del agitado interior de nuestro planeta iluminan cómo podría empezar la próxima inversión

Gary A. Glatzmaier y Peter Olson

Damos por sentado que las brújulas señalan el norte. Los marinos, a lo largo de los milenios, se han valido del campo magnético de la Tierra para navegar, como las aves y otros animales sensibles al magnetismo. Sin embargo, la orientación de los polos magnéticos del planeta no ha sido siempre la misma de hoy.

Hay minerales que registran las antiguas orientaciones del campo magnético terrestre. Revelan que ha cambiado de norte a sur y de sur a norte cientos de veces durante los 4500 millones de años que lleva existiendo nuestro planeta. Pero no ha habido cambio desde hace 780.000 años, período bastante más largo que el tiempo medio entre inversiones: unos 250.000 años. Más aún, desde que en la década de 1830 se midió por primera vez el campo geomagnético primario, ha decrecido cerca de un 10 por ciento, unas 20 veces más rápido de lo que declinaría por naturaleza si perdiese su fuente de energía. ¿No estará de camino otra inversión?

Se sabe desde hace tiempo que el fluctuante campo magnético se origina en el interior más profundo de la Tierra. Nuestro planeta, igual que otros cuerpos del sistema solar, genera su propio campo magnético mediante una dinamo interna. En líneas generales, la dinamo terrestre funciona como un generador eléctrico doméstico: crea campos eléctricos y magnéticos gracias a la energía cinética de sus partes móviles. En un generador, las partes móviles son las bobinas de alambre giratorias. En un planeta o en una estrella, se mueve un fluido eléctricamente conductor: un vasto océano de hierro fundido que quintuplica con creces el volumen de la luna, circula por el núcleo de la Tierra y constituye la geodinamo.

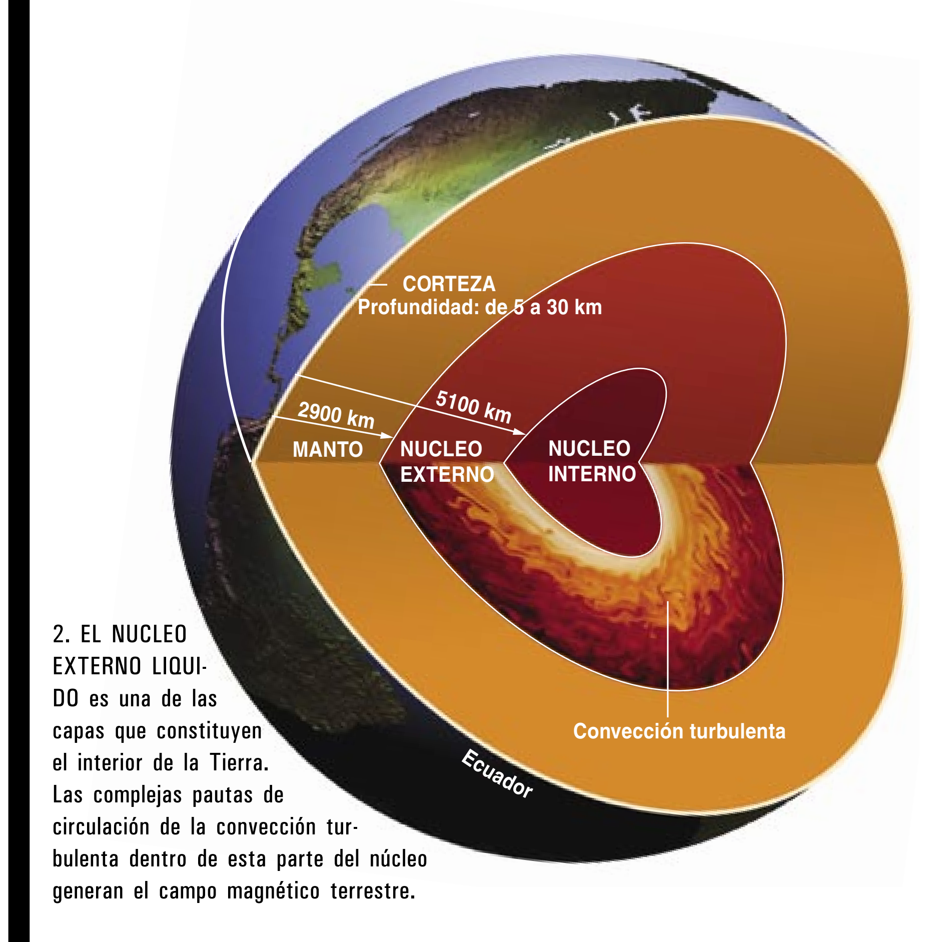
Hasta no hace mucho, la geodinamo y el magnetismo terrestre se explicaban mediante teorías simples. En los últimos diez años se han abierto nuevas vías para explorar en detalle su funcionamiento. Los satélites proporcionan claras instantáneas del campo geomagnético en la superficie terrestre. La simulación de la dinamo de la Tierra por medio de superordenadores y modelos físicos de laboratorio va esclareciendo esas observaciones orbitales. Se están explicando así las pasadas inversiones de la polarización y se columbra la manera en que se iniciaría la próxima.

El origen de la geodinamo

Antes de explorar el mecanismo de inversión del campo magnético terrestre, debe considerarse cuál es el motor de la geodinamo. Hacia la década de 1940, se sabía ya que se requerían tres condiciones básicas para que se generara un campo magnético planetario; los hallazgos posteriores se han construido sobre esa misma terna. Un gran volumen de fluido eléctricamente conductor, el núcleo externo terrestre rico en hierro líquido, es la primera de esas condiciones. Esta capa, situada bajo los 2900 kilómetros de espesor de la roca viva del manto y la fina corteza continental y oceánica, rodea el sólido núcleo interno, de hierro casi puro. El peso de la corteza y el manto crea presiones medias en el núcleo dos millones de veces superiores a las conocidas en la superficie del planeta. Las

GARY A. GLATZMAIER

1. ESTAS LINEAS DE FUERZA MAGNETICA procedentes de una simulación por ordenador de la geodinamo muestran que el campo magnético terrestre es más simple fuera de la Tierra que en el interior de su núcleo (*tubos enredados en el centro*). En la superficie del planeta, la parte principal del campo magnético sale (*tubos amarillos largos*) cerca del polo sur y entra (*tubos azules largos*) cerca del polo norte.



2. EL NUCLEO EXTERNO LIQUIDO es una de las capas que constituyen el interior de la Tierra. Las complejas pautas de circulación de la convección turbulenta dentro de esta parte del núcleo generan el campo magnético terrestre.

temperaturas del núcleo son igualmente extremas: unos 5000 grados, más o menos la temperatura de la superficie del Sol.

Estas características tan extremas del entorno aportan el segundo requisito de las dinamos planetarias: una fuente de energía para mover el fluido. La energía que impulsa la geodinamo es en parte térmica y en parte química; ambos componentes provocan ascensos de material en lo más profundo del núcleo. Igual que una cazuela de sopa hirviendo a fuego lento, el núcleo está más caliente en sus profundidades que en la parte superior (las altas temperaturas del núcleo son producto del calor atrapado en el centro de la Tierra durante su formación). Por eso, el hierro más caliente y menos denso del núcleo inferior tiende a ascender como las burbujas de una sopa caliente. Cuando el fluido alcanza el límite superior del núcleo, pierde parte de su calor, que comunica al manto que tiene encima. El hierro líquido se enfría, se vuelve más denso que el medio circundante y se hunde. Este proceso de transferencia de calor desde el fondo hasta el límite superior mediante el ascenso y descenso de fluidos recibe el nombre de convección térmica.

En la década de 1960, Stanislav Braginsky, actualmente en la Universidad de California en Los Angeles, razonó que el calor que se escapa del núcleo superior aumenta el tamaño del núcleo interior sólido, lo que produce dos fuentes más de movimiento ascendente para impulsar la convección. A medida que el hierro líquido se solidifica en cristales sobre la superficie externa del núcleo interno sólido, el calor latente se libera como subproducto. Este calor contribuye al ascenso térmico. Además, los componentes químicos menos densos, entre ellos el sulfuro de hierro y el óxido de hierro, son expulsados de los cristales del núcleo interno y ascienden hasta el núcleo externo, potenciando también la convección.

Para que en un planeta se produzca un campo magnético que se mantenga por sí solo, un tercer factor es necesario: la rotación. Mediante el efecto Coriolis, la rotación de la Tierra desvía los fluidos ascendentes del interior del núcleo terrestre; lo hace del mismo modo que desvía las corrientes oceánicas y las tormentas tropicales, a las que da la forma espiral que se les ve en las imágenes tomadas por satélites. En el núcleo, la fuerza de Coriolis da al fluido ascendente una trayectoria helicoidal, como un tirabuzón.

Que la Tierra tenga un núcleo líquido rico en hierro, la energía suficiente para impulsar la convección y una fuerza de Coriolis que desvía el fluido convectivo: éstas son las tres causas principales de que la geodinamo se haya mantenido por sí misma durante miles de millones de años. Pero no basta con saberlo para esclarecer todas las cuestiones acerca del campo generado y hallar el porqué de las inversiones.

Mapas del campo magnético

En los últimos cinco años, al poderse comparar mapas precisos del campo magnético terrestre separados por veinte años, se ha ido fraguando un gran descubrimiento. El satélite Magsat midió el campo geomagnético sobre la superficie de la Tierra en 1980; un segundo satélite, el Oersted, ha estado haciendo lo mismo desde 1999 (*véase la figura 4*). Con la premisa de que las corrientes eléctricas

Resumen/El magnetismo y la turbulencia

- La historia geológica revela que el campo magnético terrestre cambia de polaridad a menudo. Hace mucho que se quiere saber el porqué.
- Los recientes modelos por ordenador del movimiento del fluido en el núcleo terrestre fundido muestran campos magnéticos similares al terrestre, incluso con inversiones de la polaridad. Pero como el movimiento del fluido en estos modelos es mucho más simple que las turbulencias que, según se cree, existen en el interior de la Tierra, no queda clara la fiabilidad de tales resultados.
- Hoy se persigue la programación de modelos tridimensionales capaces de simular la turbulencia. Quizá despejen algún día las incertidumbres acerca de las simulaciones. Mientras, los mapas del campo magnético, basados en las observaciones por satélite, y los experimentos sobre la convección realizados en el laboratorio amplían el conocimiento del fenómeno.

JEN CHRISTIANSEN (*corte*); GARY A. GLATZMAIER (*núcleo externo*)

del manto terrestre son insignificantes, se extrapolan esas mediciones orbitales para obtener el campo en la parte superior del núcleo, en el límite núcleo-manto. No cabe acercarse más al campo (mucho más intenso y complejo) del interior del núcleo, el lugar donde de verdad se originan las fluctuaciones magnéticas: las fuertes corrientes eléctricas del núcleo impiden la medición directa de su campo magnético. A pesar de las limitaciones inherentes, se han realizado así observaciones muy interesantes, que han registrado indicios del posible comienzo de una nueva inversión de la polaridad.

Así, se sabe ahora que la mayor parte del campo magnético terrestre se origina en cuatro extensas regiones del límite núcleo-manto. Aunque la geodinamo produce un intenso campo magnético, sólo un 1 por ciento de su energía se extiende más allá del núcleo. La estructura dominante de ese campo, cuando se mide en la superficie, es la dipolar. La mayor parte del tiempo está casi alineada con el eje de rotación de la Tierra. Igual que en un simple imán, este flujo magnético primario sale del núcleo por el hemisferio sur y entra de nuevo en él por el hemisferio norte (las brújulas señalan el polo norte geográfico porque en su vecindad se encuentra el polo magnético sur del dipolo). Sin embargo, los satélites revelaron que el flujo no se distribuye con uniformidad por todo el globo. La mayor parte de la intensidad total del dipolo se origina bajo Norteamérica, Siberia y la costa de la Antártida.

Ulrich R. Christensen, del Instituto Planck para la Investigación del Sistema Solar, en Katlenburg-Lindau, sospecha que estas grandes zonas van y vienen durante miles de años y son producto de la convección nuclear en continua evolución. ¿Puede un fenómeno similar ser la causa de las inversiones del dipolo? El registro geológico nos informa de que las inversiones del pasado acontecieron a lo largo de períodos más bien cortos, de unos 4000 a 10.000 años. El dipolo tardaría cerca de 100.000 años en desaparecer por sí mismo en caso de que se detuviera la geodinamo. Una transición tan rápida implica que algún tipo de inestabilidad destruye la polaridad original, al tiempo que genera una nueva polaridad.

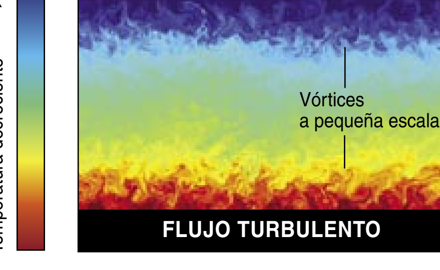
En el caso de las inversiones individuales, esa misteriosa inestabilidad consiste probablemente en algún tipo de cambio caótico en la estructura del flujo que sólo de vez en cuando consiga la inversión del dipolo global. Pero la frecuencia de las inversiones, que se ha incrementado a ritmo constante en los últimos 120 millones de años (*véase la figura 1*), podría estar sujeta a un mecanismo de control externo, quizás un cambio de temperatura en el manto más profundo que fuerce al núcleo a cambiar sus pautas de ascenso de material.

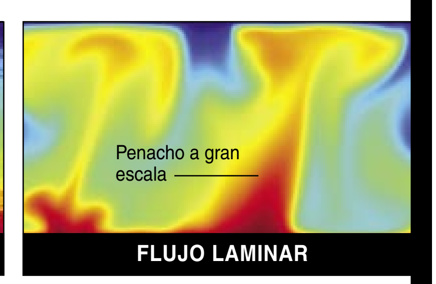
Cuando otro grupo analizó los mapas de los satélites Magsat y Oersted, encontró síntomas de un posible cambio inductor de inversiones. Gauthier Hulot y sus compañeros del Instituto Geofísico de París descubrieron que las variaciones duraderas del campo magnético proceden de lugares del límite núcleo-manto en los que la dirección del flujo es *opuesta* a la normal en el hemisferio de que se trate. La mayor de esas zonas de flujo invertido se extiende desde la parte que cae bajo el extremo sur del Africa occidental hasta la que subyace bajo el extremo meridional de Sudamérica. En esta zona invertida, el flujo magnético apunta hacia dentro, hacia el núcleo, cuando la mayor parte del flujo en el hemisferio sur apunta hacia fuera.

Creación de zonas

Al compararse las recientes medidas magnéticas del Oersted con las de 1980, se llegó también a la importante conclusión de que siguen formándose nuevas zonas de flujo invertido en el límite núcleo-manto. Así ocurre, por ejemplo, bajo la costa este de Norteamérica y en el Artico. Aún más, las zonas más antiguas han aumentado y se han desplazado ligeramente hacia los polos. A finales de la década de 1980, David Gubbins, de la Universidad de Leeds, con mapas viejos y rudimentarios del campo magnético, se percató de que la proliferación, crecimiento y migración hacia los polos de estas zonas de flujo invertido daba cuenta del declive histórico del dipolo.

Estas observaciones pueden explicarse físicamente mediante una abstracción, las líneas magnéticas de fuerza (en realidad, el campo es continuo en el espacio). Podemos considerar que esas líneas de fuerza están "congeladas" en el núcleo de hierro fluido y, por consiguiente, que su movimiento es solidario con el del hierro, como un hilo de tinte que se retuerce en un vaso de agua al remover ésta. En el núcleo terrestre, en virtud del efecto de Coriolis, los vórtices del fluido distorsionan las líneas magnéticas de fuerza hasta disponerlas en haces de tubos. Cada torsión aprisiona más líneas de fuerza en el núcleo e incrementa la energía del campo magnético. Si este proceso continuara sin límites, el campo magnético crecería indefinidamente. Sin embargo, la resistencia eléctrica tiende a difuminar y suavizar los giros de las líneas del campo magnético, hasta el punto de frenar la expansión del campo magnético sin eliminar la dinamo.





3. EL COMPLEJO COMPORTAMIENTO DE LOS FLUJOS en el fundido núcleo externo terrestre se asemeja al desarrollado en las simulaciones bidimensionales de la convección turbulenta (*izquierda*). En las simulaciones de la dinamo en tres dimensiones, sin embargo, la falta de potencia de cálculo obliga a limitarse a estudiar los penachos del flujo laminar, que suelen ser mayores (*derecha*) que en los modelos de dos dimensiones (recuerdan al aceite mineral caliente que asciende en una "lámpara de lava"). Por el momento, los ordenadores están muy lejos de resolver los complejos cálculos asociados al flujo turbulento tridimensional del núcleo terrestre.

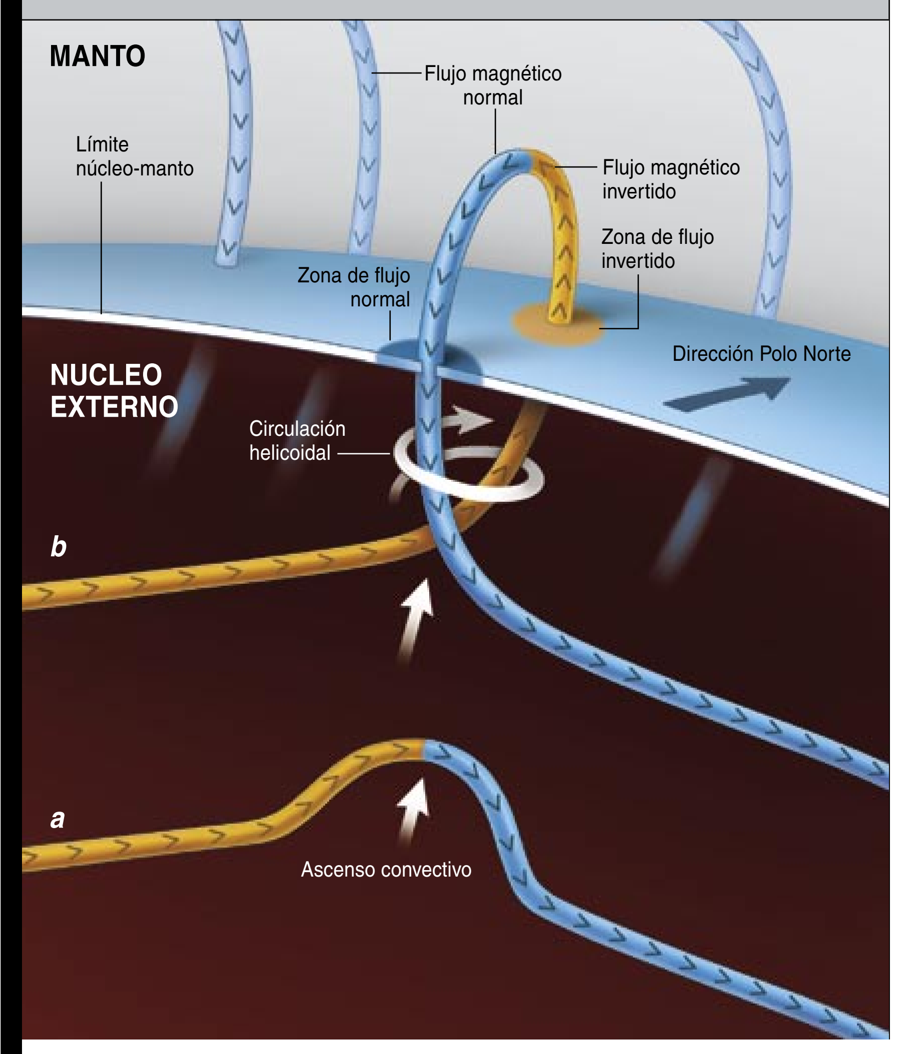
GARY A. GLATZMAIER

En el límite núcleo-manto se forman zonas de flujo magnético intenso, sea normal o invertido, cuando los remolinos y los vórtices interactúan con los campos magnéticos toroidales de dirección este-oeste inmersos en el núcleo. Esos movimientos turbulentos del fluido pueden curvar y torcer las líneas del campo toroidal hasta configurarlas como lazos, los campos poloidales, de orientación norte-sur. Algunas veces, es un fluido que asciende el que provoca la curvatura. Si ese movimiento ascendente tiene fuerza suficiente, la parte más alta del lazo poloidal saldrá del núcleo (*véase el recuadro* "Zonas de flujo invertido"). Esta expulsión crea dos zonas de flujo, allí donde el borde del lazo atraviesa el límite núcleo-manto. Una de las zonas tiene flujo normal (en la misma dirección que el campo dipolar en ese hemisferio); la otra, lo tendrá en el sentido opuesto: es de flujo invertido.

Cuando la desviación hace que la zona de flujo invertido se acerque más al polo geográfico que la zona de flujo normal, el resultado es un debilitamiento del dipolo, que cerca de sus polos es más sensible a los cambios. En estos momentos, se está dando una situación así en la zona de flujo invertido que se halla bajo el extremo meridional de Africa. Para que se produzca una inversión de la polaridad a escala planetaria, la zona de flujo invertido debería crecer y engullir entera la región polar; al mismo tiempo, debería ocurrir, en el otro polo geográfico, un cambio similar en la polaridad magnética global de la región.

ZONAS DE FLUJO INVERTIDO

Cuando las retorcidas líneas del campo magnético atraviesan el límite superior del núcleo terrestre y llegan al manto, nacen regiones en las que el flujo magnético discurre en dirección opuesta a la común en el hemisferio de que se trate. Estas zonas de flujo invertido atenúan la componente principal del campo magnético de la superficie, la dipolar, y quizá hasta anuncien una inversión global de la polaridad. Las zonas de flujo invertido se originan cuando el fluido que asciende por el núcleo externo fundido empuja hacia arriba las líneas casi horizontales del campo magnético del núcleo. A veces, este ascenso convectivo empuja una línea hasta el punto de dibujar una joroba (*a*). A la vez, la rotación de la Tierra impulsa la circulación helicoidal del fluido fundido, que dobla la joroba hasta que se convierte en un lazo (*b*). Cuando la fuerza ascendente es lo bastante potente para expulsar el lazo del núcleo, se crean dos zonas de flujo opuesto en el límite núcleo-manto.



JEN CHRISTIANSEN

Simulaciones por superordenador

Se simula la geodinamo en superordenadores y en el laboratorio. Así se quiere conocer mejor las zonas de flujo invertido y discernir las señales con que podrían anunciarnos la próxima inversión de la polaridad. La era moderna de las simulaciones informáticas de la dinamo empezó en 1995, cuando tres grupos —el de Akira Kageyama, de la Universidad de Tokio; el de Paul H. Roberts, de la Universidad de California en Los Angeles, y uno de nosotros (Glatzmaier); y el de Christopher A. Jones, de la Universidad de Exeter— produjeron, cada uno por su parte, simulaciones numéricas que predecían campos magnéticos parecidos al de la superficie terrestre. Desde entonces, las simulaciones, que abarcan períodos de cientos de miles de años, han demostrado que la convección genera en el límite núcleo-manto zonas de flujo invertido del estilo de las observadas en las imágenes de satélite. Estas zonas aparecen a menudo antes de una inversión espontánea del dipolo magnético, mutación global que algunas simulaciones pueden también reproducir.

Gracias a las inversiones de la polaridad generadas por ordenador, se vislumbró por fin cómo se originaban y progresaban en la naturaleza (*véase el recuadro* "Simulaciones de las inversiones de la polaridad"). En una simulación tridimensional —que para reproducir 300.000 años tuvo que ejecutarse durante 12 horas de

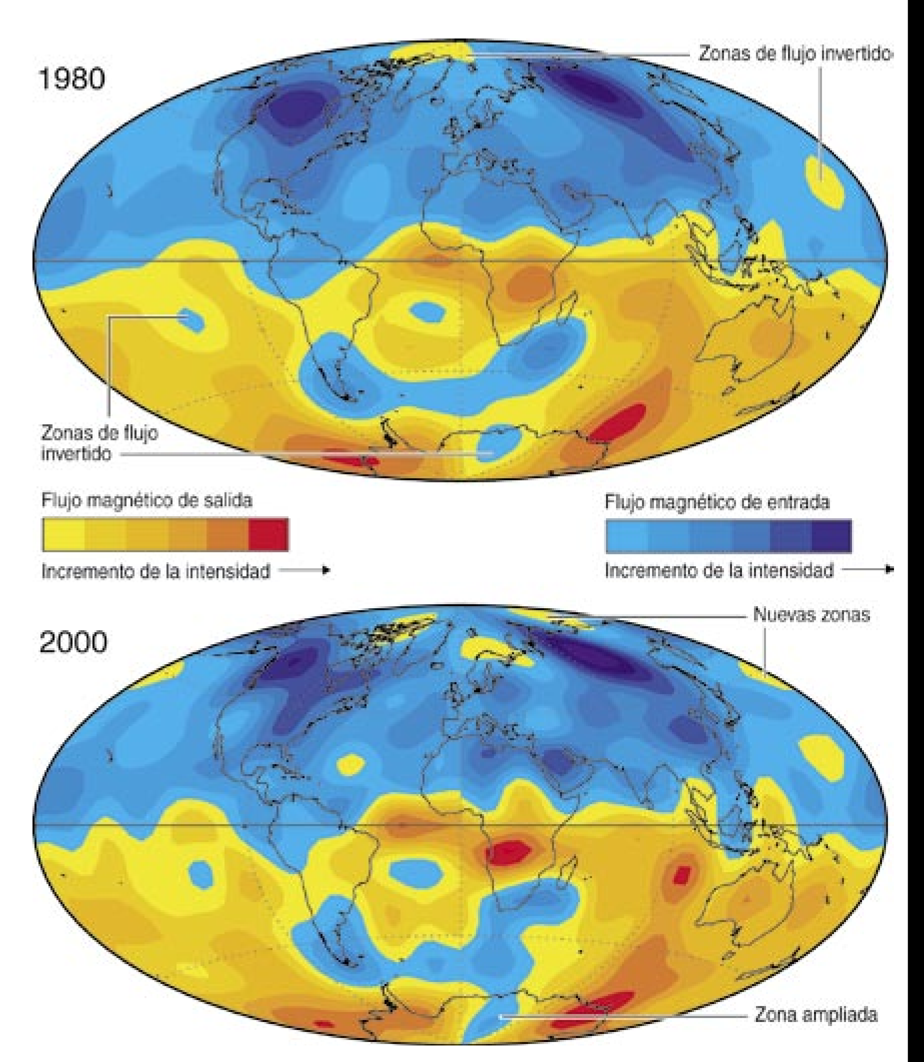
cada día a lo largo de más de un año—, el inicio de una inversión se ligó al decrecimiento de la intensidad del campo dipolar. Aparecieron a continuación varias zonas de flujo magnético invertido, como las que actualmente se forman en el límite núcleo-manto. Pero en vez de extinguir el campo magnético, las zonas de flujo invertido crearon durante la transición un débil campo con una compleja mezcla de polaridades.

En la superficie de ese modelo de la Tierra, la inversión del dipolo se produjo cuando las zonas de flujo invertido dominaron sobre la polaridad original del límite núcleo-manto. En total, se necesitaron 9000 años para la disipación de la antigua polaridad y para que la nueva se extendiera a todo el núcleo.

Lo que falta

En parte gracias a estos éxitos, asistiremos a una proliferación de modelos informáticos de la dinamo. Según las últimas informaciones, más de una docena de grupos de distintas partes del mundo los emplean para estudiar los campos magnéticos de los cuerpos del sistema solar e incluso de astros más lejanos. Pero, ¿en qué medida los modelos de la geodinamo representan con fidelidad la dinamo terrestre existente? No se sabe.

Por el momento, ningún modelo informático de la dinamo ha reproducido el amplio espectro de la turbulencia existente en el interior del planeta. Los superordenadores en paralelo no son aún tan veloces como para simular de manera precisa, en tres dimensiones y con parámetros físicos realistas la turbulencia magnética. Cabe suponer que los remolinos y vórtices turbulentos de menor tamaño que desvían el campo magnético en el núcleo terrestre actúan a una escala de metros a decenas de metros, demasiado pequeña para los actuales modelos globales de la geodinamo ejecutados por superordenadores. De ahí que los modelos informáticos tridimensionales de la geodinamo sólo hayan conseguido simular el flujo simple y a gran escala de la convección laminar, semejante al movimiento del aceite mineral caliente de una de esas "lámparas de lava" que se pueden ver en las tiendas de artículos de regalo.



4. LOS MAPAS DE CONTORNO del campo magnético terrestre, extrapolados al límite núcleo-manto a partir de mediciones por satélite, indican que en el hemisferio sur la mayor parte del flujo magnético sale del núcleo, mientras que en el hemisferio norte entra en él. Pero hay ciertas regiones donde ocurre al revés. Estas zonas de flujo invertido proliferaron y se ampliaron entre 1980 y 2000; si alcanzaran los polos, se produciría posiblemente una inversión de la polaridad.

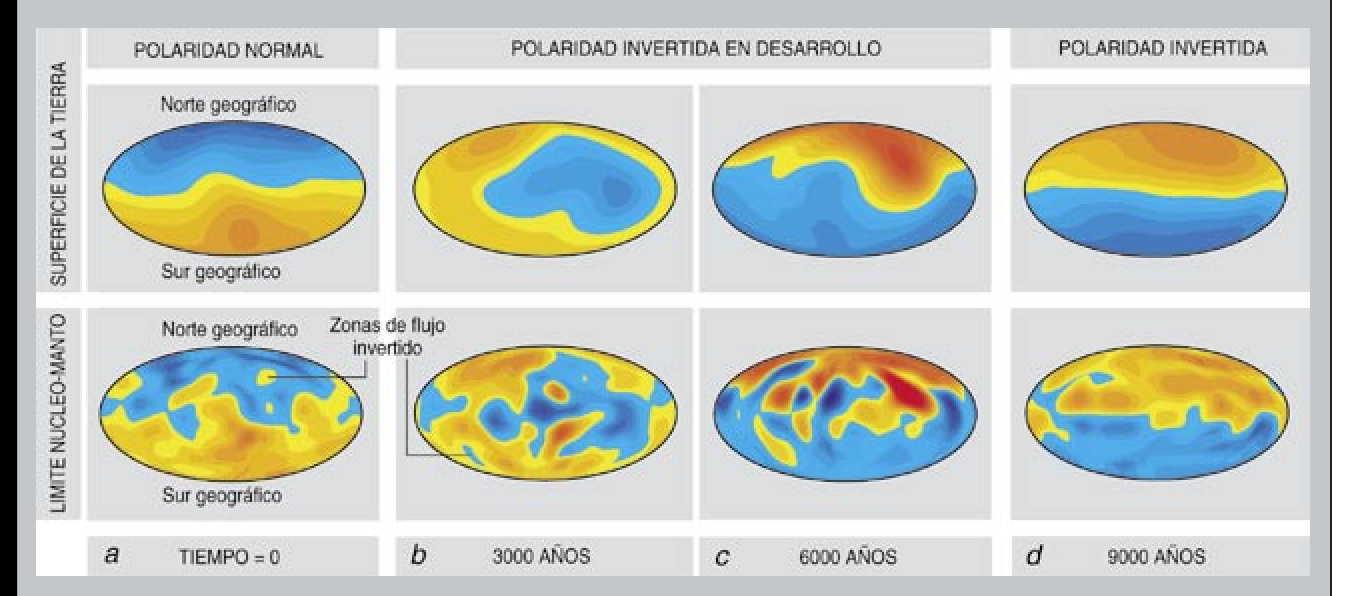
PETER OLSON; FUENTE: GAUTHIER HULOT ET AL. EN *NATURE*, VOL. 416, PAGS. 620-623; 11 DE ABRIL DE 2002

Para reproducir con cierto grado de aproximación los efectos del flujo turbulento en los modelos laminares, se pueden asignar a ciertas propiedades del núcleo fluido valores mayores de los que tienen en la realidad; su pequeñez impide abordarlas numéricamente tal y como son. En aras de conseguir una turbulencia realista mediante un modelo informático, hay que limitarse a una versión bidimensional, pero lo que se gana por ahí se pierde por otro lado: un flujo en dos dimensiones no puede mantener una dinamo. Con todo, incluso estos modelos restringidos dan ya a entender que los flujos laminares generados en las actuales simulaciones de la geodinamo son más tranquilos y simples que los flujos turbulentos que casi con toda seguridad existen en el núcleo terrestre.

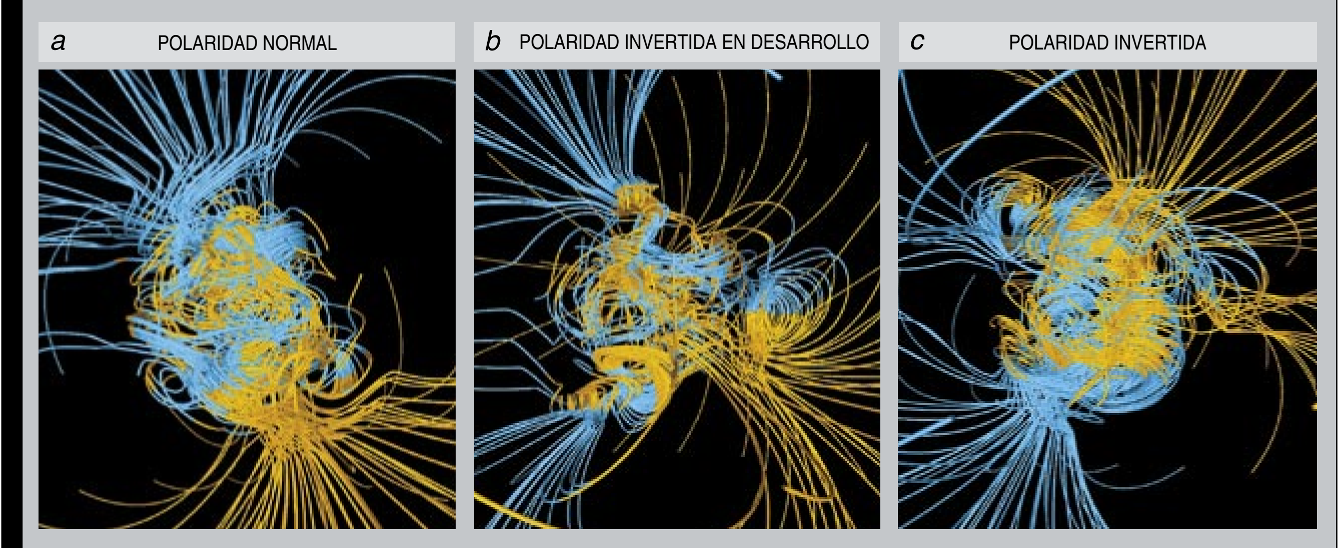
Es probable que la diferencia más significativa radique en los caminos que sigue el fluido en su ascenso por el núcleo. En las meras simulaciones de la convección laminar, se extienden grandes penachos desde el fondo del núcleo hasta su límite superior. En cambio, en los turbulentos modelos bidimensionales la convección contiene múltiples penachos y vórtices de pequeña escala que se desprenden cerca de los límites superior e inferior del núcleo e interaccionan dentro de la zona principal de convección, situada entre ambos límites.

SIMULACIONES DE LAS INVERSIONES DE LA POLARIDAD

Las simulaciones tridimensionales por ordenador de la geodinamo consiguen ya generar inversiones espontáneas del dipolo magnético terrestre. Arrojan así luz sobre el origen de las inversiones conservadas en el registro geológico (*véase la escala temporal de la página siguiente*). Una de las inversiones generadas por el modelo del que fue coautor uno de los autores (Glatzmaier) tardó 9000 años en completarse. El proceso queda descrito en estos mapas de la parte vertical del campo magnético de la superficie terrestre y del límite núcleo-manto, donde el campo es más complejo. Los modelos que emplean líneas de campo magnético proporcionan una tercera opción para visualizar la inversión de la polaridad.

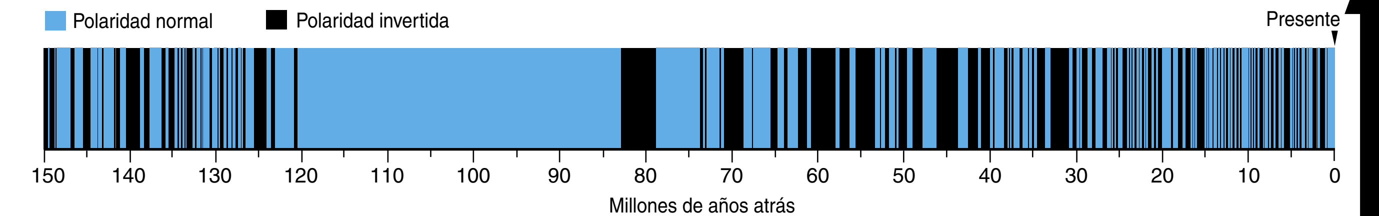


LOS MAPAS DEL CAMPO MAGNETICO parten de una polaridad normal: la mayor parte del flujo magnético planetario sale del núcleo (*amarillo*) en el hemisferio sur y entra en el núcleo (*azul*) por el hemisferio norte (*a*). Anuncian el comienzo de la inversión distintas áreas de flujo magnético invertido (*azul en el sur y amarillo en el norte*); recuerdan a las zonas de flujo invertido que se están formando ahora en el límite núcleo-manto. En unos 3000 años, las zonas de flujo invertido atenúan la intensidad del campo dipolar hasta que lo sustituye un campo de transición, más débil aunque más complejo, en el límite núcleo-manto. (*b*). A los 6000 años, las zonas de flujo invertido dominan a la polaridad original del límite núcleo-manto (*c*). Observando sólo la superficie, la inversión parece ya completa en este momento, aunque se requieren otros 3000 años para que el dipolo del núcleo se invierta totalmente (*c*).



EL MODELO representa el campo magnético del núcleo (*líneas enmarañadas del centro*) y el dipolo que aflora al exterior (*líneas largas y curvas*) 500 años antes de que la inversión del dipolo magnético llegue a su mitad (*a*), a la mitad del proceso (*b*) y 500 años después de la mitad, cuando la inversión casi se ha completado (*c*).

GARY A. GLATZMAIER; SIMULADO EN EL CENTRO DE SUPERCOMPUTACION DE PITTSBURGH



5. SE HAN SUCEDIDO cientos de inversiones de la polaridad en los últimos 150 millones de años, con distintas frecuencias; también debía de ocurrir así antes. Esas inversiones se descubrieron en los minerales magnéticos, cuya imanación se alinea con el campo magnético terrestre cuando se calientan y mantiene esa orientación cuando se enfrían.

JEN CHRISTIANSEN; FUENTE: D. V. KENT Y W. LOWRIE EN *TIMESCALES OF THE PALEOMAGNEITC FIELD*. DIRIGIDO POR J.E.T. CHANNELL ET AL; AGU, 2004

Tales diferencias en la distribución del flujo podrían influir mucho en la estructura del campo magnético terrestre y en el tiempo necesario para que se den cambios. De ahí que se esté poniendo mucho empeño en crear una nueva generación de modelos tridimensionales. Algún día, quizá dentro de diez años, los avances en la velocidad de los procesadores informáticos permitirán realizar potentes simulaciones de la dinamo turbulenta. Mientras, esperamos aprender más sobre la dinamo gracias a experimentos de laboratorio.

Dinamos de laboratorio

Comparar las dinamos informáticas (que carecen de turbulencia) con las dinamos de laboratorio (que carecen de convección) resulta útil para mejorar nuestro conocimiento de la geodinamo. La viabilidad de las dinamos de laboratorio se demostró ya en la década de los sesenta, pero el camino hacia el éxito todavía sería largo. La enorme diferencia entre el tamaño de los aparatos de laboratorio y el del núcleo de un planeta suponía un factor de vital importancia. Para que una dinamo fluida se mantenga por sí misma se requiere que cierto parámetro adimensional, el número magnético de Reynolds, supere un valor mínimo de aproximadamente 10.

El núcleo terrestre tiene un número magnético de Reynolds muy grande, alrededor de 1000, a causa sobre todo de su gran dimensión lineal (el radio del núcleo es de unos 3485 kilómetros). Resulta dificilísimo crear un número magnético de Reynolds elevado en pequeños volúmenes de fluido, a no ser que éste se mueva a velocidades enormes.

El viejo sueño de generar un campo magnético espontáneo en una dinamo fluida de laboratorio se realizó en el año 2000. Dos grupos europeos —el de Agris Gailitis, de la Universidad de Letonia, y el de Robert Stieglitz y Ulrich Müller, del Centro de Investigaciones de Karlsruhe, y Fritz Busse, de la Universidad de Bayreuth— lograron por separado la autogeneración en grandes volúmenes de sodio líquido (se utilizó el sodio líquido por su alta conductividad eléctrica y su bajo punto de fusión). Ambos grupos crearon flujos de alta velocidad con sistemas de tubos helicoidales de uno a dos metros de longitud. Obtuvieron así un número magnético crítico aproximado de 10 [*véase* "La geodinámica en el laboratorio", por Ulrich Müller y Robert Stieglitz; INVESTIGACIÓN Y CIENCIA, mayo de 2002].

Estos experimentos respaldaron la teoría de la dinamo. Nos da cierta confianza de que acertamos al aplicarla a la Tierra y otros planetas. En la actualidad, muchos grupos están desarrollando la próxima generación de dinamos de laboratorio. Para simular mejor la geometría de la Tierra, esos experimentos agitarán el sodio líquido en cámaras esféricas, la mayor de ellas de tres metros de diámetro.

Aparte de los proyectos en marcha para conseguir dinamos de laboratorio y simulaciones informáticas tridimensionales más realistas, el satélite internacional CHAMP está cartografiando el campo geomagnético con precisión suficiente para medir en tiempo real los cambios que se producen en el límite núcleo-manto. Se prevé que este satélite proporcione una imagen ininterrumpida del campo geomagnético a lo largo de los cinco años que durará su misión. Así, se mantendrán bajo observación el crecimiento continuo de las zonas de flujo invertido y demás indicios de desvanecimiento del campo dipolar.

Anticipamos que dentro de diez o veinte años se producirá una síntesis de los tres métodos —las observaciones por satélite, las simulaciones por ordenador y los experimentos de laboratorio—. Con una visión más completa de la geodinamo, sabremos si las ideas actuales sobre el campo magnético y sus inversiones van por el buen camino.

Los autores

Gary A. Glatzmaier y Peter Olson elaboran modelos informáticos para estudiar la estructura y la dinámica del interior de los planetas y las estrellas. A mediados de los años noventa, Glatzmaier, entonces en el Laboratorio Nacional de Los Alamos, creó (junto a Paul H. Roberts, de la Universidad de California en Los Angeles) la primera simulación de la geodinamo que produjo una inversión magnética espontánea del dipolo. Enseña desde 1998 en el departamento de ciencias de la Tierra de la Universidad de California en Santa Cruz. Olson ingresó en el departamento de ciencias planetarias y de la Tierra de la Universidad Johns Hopkins en 1978.

Bibliografía complementaria

NUMERICAL MODELLING OF THE GEODYNAMO: MECHANISMS OF FIELD GENERATION AND EQUILIBRATION. Peter Olson, Ulrich Christensen y Gary A. Glatzmaier en *Journal of Geophysical Research*, vol. 104, n.º B5, páginas 10.383-10.404; 1999.

EARTH'S CORE AND THE GEODYNAMO. Bruce A. Buffet en *Science*, vol. 288, páginas 2007-2012; 16 de junio de 2000.

GEODYNAMO SIMULATIONS: HOW REALISTIC ARE THEY? Gary A. Glatzmaier en *Annual Review of Earth and Planetary Sciences*, vol. 30, páginas 237-257; 2002.

RECENT GEODYNAMO SIMULATIONS AND OBSERVATIONS OF THE GEOMAGNETIC FIELD. Masaru Kono y Paul H. Roberts en *Reviews of Geophysics*, vol. 40, n.º 4, página 1013; 2002.

El otro genoma

El viejo axioma “un gen, una proteína” ha dejado de tener vigencia. Cuanto más complejo es un organismo, tanta mayor probabilidad hay de obtener múltiples proteínas a partir de un solo gen

Gil Ast

1. RATONES Y HUMANOS comparten el 88 por ciento de su genoma. Gran parte de las diferencias entre humanos y roedores se originan en el proceso de edición de la información génica.



BILL STEELE

La primavera del 2000 sorprendió a los biólogos moleculares haciendo apuestas sobre el número de genes que se descubrirían en el genoma humano cuando acabara la secuenciación de sus nucleótidos. Había quien elevaba la cifra a 153.000. Después de todo, se comentaba, si el hombre fabrica unos 90.000 tipos diferentes de proteínas, se contaría con el mismo número de genes para codificarlas, por lo menos. Además, dada nuestra complejidad, portaríamos una dotación genética mucho mayor que la del nemátodo *Caenorhabditis elegans*, que tiene 1000 células y un genoma de 19.500 genes, o el maíz, con 40.000 genes.

Cuando, en verano de aquel mismo año, se publicó el primer borrador del genoma humano, más de uno se quedó perplejo por la cifra calculada: de 30.000 a 35.000 genes codificadores de proteínas. Una cuantía casi vergonzante. A medida que progresó la cartografía del genoma humano, la revisión del número exacto de genes fue a la baja: por debajo incluso de los 25.000. Al propio tiempo, los expertos han venido perfilando una nueva interpretación de los resultados: nuestro bajo recuento podría considerarse una señal de complejidad; el hombre hace un uso sumamente versátil de esa gavilla insignificante de genes.

A través de un mecanismo de corte y empalme alternativo, la información almacenada en los genes de organismos complejos se somete a un proceso de edición harto flexible, de suerte tal que un solo gen especifique dos o más proteínas distintas. La comparación del genoma humano con el de otros organismos ha puesto de manifiesto que esa edición alternativa —llamemos así a la flexibilidad del proceso de corte y empalme— explica la diversidad existente entre organismos que portan una dotación genética bastante similar. En un individuo, este mecanismo versátil permite que los tejidos diferentes acometan funciones diversas a partir de un mismo surtido de genes, harto exiguo.

El predominio de la edición alternativa aumenta con la complejidad de un organismo. En el hombre, se hallan sujetos a edición alternativa hasta tres cuartos del total de sus genes. Con toda probabilidad, el propio mecanismo contribuyó a la evolución de la complejidad y podría impulsar también nuestra evolución futura. De momento, se empiezan ya a comprender los mecanismos en cuya virtud una edición defectuosa puede desembocar en diversos tipos tumorales y enfermedades congénitas. Se barrunta también una posible aplicación terapéutica del proceso de corte y empalme.

Intrones y exones

La edición alternativa reviste suma importancia para el desenvolvimiento normal de los organismos. De ella dependen la vida y la muerte de una célula dañada. Cada célula percibe constantemente las condiciones de su exterior e interior, para así decidir si persiste en su desarrollo o se autoinmola en un proceso de apoptosis. Las células que no pueden reparar su ADN activan el programa apoptótico. Craig B. Thompson y su grupo, de la Universidad de Pennsylvania, acaban de demostrar que el gen *Bcl-x*, regulador de la apoptosis, se somete a una edición alternativa para producir o una proteína Bcl-x(L) o una proteína Bcl-x(S). La primera acalla la apoptosis; la segunda la promueve.

El descubrimiento de la capacidad celular de originar formas proteicas distintas a partir de un solo gen se produjo hace unos 25 años; pero tal posibilidad se reputó una rareza. De la genómica comparada reciente, por contra, se desprende que no sólo se trata de un proceso habitual, sino también crucial. Recibe así un nuevo giro, radical, la interpretación de la traducción de la información génica en proteína.

Siguen vigentes la mayoría de los trazos divulgados de la doctrina: el genoma contiene todas las instrucciones necesarias para fabricar y mantener un organismo, codificadas en un lenguaje de cuatro letras que corresponden a nucleótidos de ADN (en forma abreviada, A, G, C y T). En los cromosomas humanos, hay unos tres mil millones de nucleótidos alineados en cada una de las dos cadenas complementarias que forman la doble hélice. Cuando llega el momento de "expresar" un gen, la cremallera de doble hebra de ADN se abre sólo durante el tiempo justo para que se fabrique, a partir de ARN (una molécula químicamente emparentada), una copia de cadena simple de la secuencia del gen.

Cada secuencia de nucleótidos de ADN que se transcribe así en una versión de ARN corresponde a un gen. Algunos tramos del transcrito primario no se traducen nunca en proteínas, sino que llevan a cabo funciones de mantenimiento y regulación en el interior de la célula. Los transcritos de ARN de genes que sí codifican proteínas serán leídos por la maquinaria celular y traducidos a la correspondiente secuencia de aminoácidos. Antes, sin embargo, el transcrito primario debe someterse a un proceso de edición.

En 1977, Phillip A. Sharp, del Instituto de Tecnología de Massachusetts, y Richard J. Roberts, de New England Biolabs, descubrieron que esos transcritos de ARN iniciales, o primarios, se parecen a libros que contienen capítulos carentes de

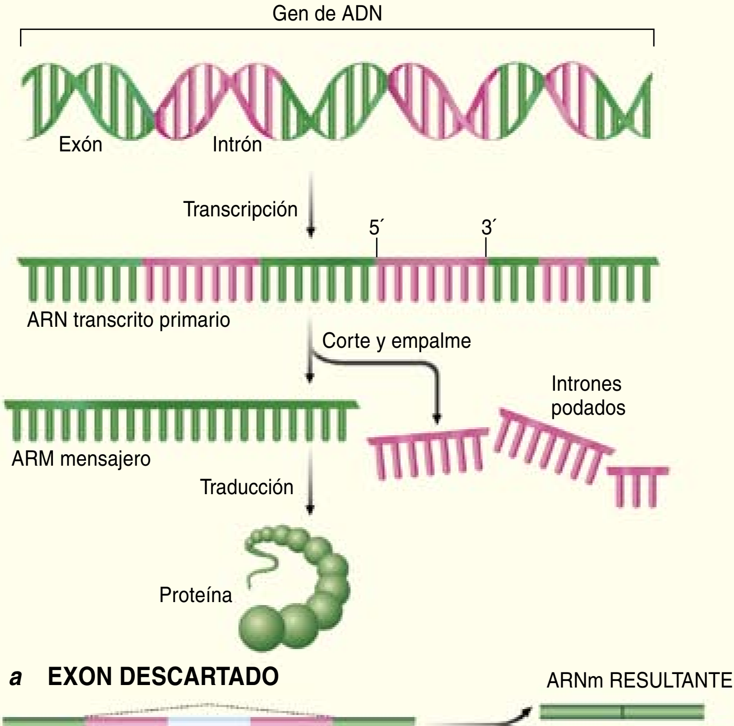
Resumen/Cortar y empalmar

- Cuando edita la información de un gen, la maquinaria celular puede transmitir múltiples significados. Basta, pues, un acervo génico limitado para generar un vasto repertorio de proteínas.
- Esta edición alternativa se conocía desde hace largo tiempo. Sin embargo, hasta que no se compararon los genomas de varias especies, no se descubrió que se trata de un mecanismo común de los organismos complejos. Contribuye, de un modo determinante, a la diversidad entre individuos con dotaciones génicas similares.
- El corte y empalme alternativo permite que un número restringido de genes produzca y mantenga organismos complejos, determinando cuándo, dónde y qué tipo de proteínas se fabrican. Quizá pronto logremos regular la edición de nuestros propios genes para combatir las enfermedades.

UN GEN, VARIAS PROTEINAS

Según la visión tradicional de la actividad génica, la expresión de un gen de ADN comienza cuando éste se transcribe en forma de ARN; luego, la maquinaria celular de edición poda los tramos redundantes (intrones) y empalma las porciones con significado (exones) para formar una versión de ARN mensajero (ARNm), que se traduce en proteína. Pero no siempre ocurre así. En los organismos complejos, el transcrito inicial de ARN se edita de forma alternativa —se descartan exones y se retienen intrones, o segmentos de ellos— para producir ARNm diversos y, por ende, proteínas distintas, a partir de un solo gen.

EXPRESION GENICA TRADICIONAL
Una secuencia de ADN se transcribe en una copia unicatenaria de ARN. La maquinaria celular "edita" entonces este transcrito primario: los intrones —cada uno definido por una señal distintiva al principio y otra al final de la secuencia de nucleótidos, los puntos de corte 5' y 3'— se podan, mientras que se ensamblan los exones para formar el ARNm. Por fin, el ARNm se traduce en una proteína.



CORTE Y EMPALME ALTERNATIVO
El transcrito primario de un gen puede ser editado de varias formas, como se muestra a la derecha (las distintas opciones de corte y empalme se indican con líneas discontinuas). Un exón puede descartarse (*a*). La maquinaria de corte y empalme puede reconocer puntos de corte 5' alternativos para un intrón (*b*) o bien puntos de corte 3' alternativos (*c*). Un intrón puede conservarse en el ARNm transcrito (*d*). Y un exón pudiera, por fin, retenerse según un criterio de exclusión mutua (*e*).

Exón que siempre se conserva
Exón editado de forma alternativa
Intrón

ARNm RESULTANTE

a **EXON DESCARTADO**

b **PUNTOS DE CORTE 5´ ALTERNATIVOS**

c **PUNTOS DE CORTE 3´ ALTERNATIVOS**

d **INTRON CONSERVADO**

e **RETENCION DE EXONES MUTUAMENTE EXCLUYENTES**

LUCY READING-IKKANDA

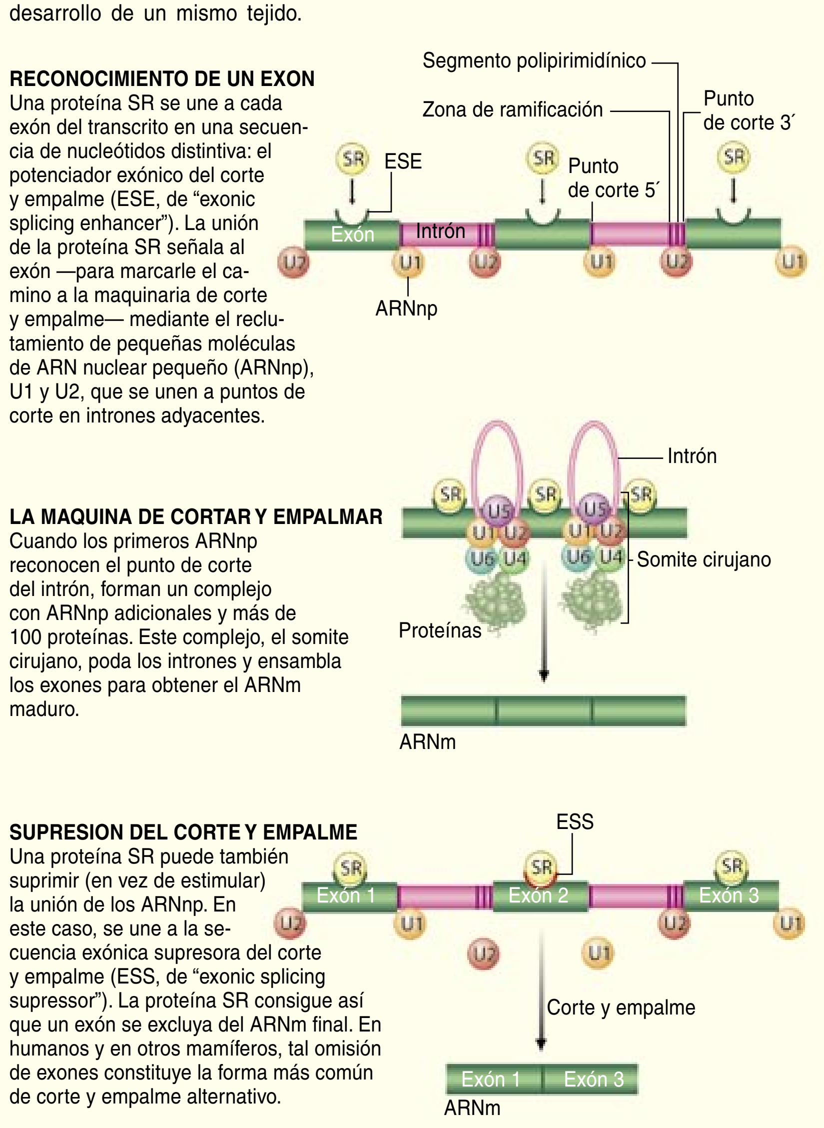
significado insertados, a intervalos, en el texto. Para que el ARN relate una historia coherente, hay que cortar los fragmentos carentes de sentido, o intrones, mientras que deben ensamblarse los capítulos con significado, o exones.

En el proceso de edición, es decir, de corte y ensamblaje, los intrones son podados del transcrito primario, en tanto que se engarzan los exones (codificadores de proteínas) para formar una versión madura del transcrito: el ARN mensajero (ARNm).

Pero en 1980, Randolph Wall, de la Universidad de California en Los Angeles, mostró que no siempre se cumplía ese modelo básico de corte y empalme del pre-ARNm, en cuyo marco los intrones se descartan y los exones se incluyen en el ARNm. En realidad, la maquinaria celular tiene la capacidad de "decidir" si cercena un exón o amnistía a un intrón —o segmentos del mismo— para el transcrito final de ARNm. Esta capacidad para editar de forma alternativa transcritos de pre-ARNm potencia la versatilidad de cualquier gen. Asimismo, confiere al mecanismo de

LOS SOMITES CIRUJANOS

Una vez que se ha creado un transcrito primario de un gen, los somites cirujanos ("spliceosome") llevan a cabo la edición del ARN. En los organismos complejos, este proceso está controlado por proteínas reguladoras del corte y empalme que reconocen a los exones y dirigen a los somites cirujanos a puntos de corte específicos. Estas moléculas reguladoras, o proteínas SR, determinan cuándo y cómo se produce la edición alternativa de un gen. Las propias proteínas SR se fabrican en diversas versiones a tenor de los tejidos y tipos celulares o según la fase de desarrollo de un mismo tejido.



RECONOCIMIENTO DE UN EXON
Una proteína SR se une a cada exón del transcrito en una secuencia de nucleótidos distintiva: el potenciador exónico del corte y empalme (ESE, de "exonic splicing enhancer"). La unión de la proteína SR señala al exón —para marcarle el camino a la maquinaria de corte y empalme— mediante el reclutamiento de pequeñas moléculas de ARN nuclear pequeño (ARNnp), U1 y U2, que se unen a puntos de corte en intrones adyacentes.

LA MAQUINA DE CORTAR Y EMPALMAR
Cuando los primeros ARNnp reconocen el punto de corte del intrón, forman un complejo con ARNnp adicionales y más de 100 proteínas. Este complejo, el somite cirujano, poda los intrones y ensambla los exones para obtener el ARNm maduro.

SUPRESION DEL CORTE Y EMPALME
Una proteína SR puede también suprimir (en vez de estimular) la unión de los ARNnp. En este caso, se une a la secuencia exónica supresora del corte y empalme (ESS, de "exonic splicing supressor"). La proteína SR consigue así que un exón se excluya del ARNm final. En humanos y en otros mamíferos, tal omisión de exones constituye la forma más común de corte y empalme alternativo.

corte y empalme un tremendo poder para determinar qué cantidad de una proteína concreta se fabrica respecto de otras proteínas codificadas por el mismo gen.

En 1984, Tom Maniatis, Michael Green y sus colaboradores de la Universidad de Harvard desarrollaron un procedimiento *in vitro* para estudiar la maquinaria molecular que lleva a cabo el corte de los intrones y el empalme de los exones. Los detalles de cómo opera, así como de su regulación, todavía no se conocen. Con todo, esta investigación está arrojando luz sobre un sistema exquisitamente intrincado y fascinante.

La máquina de cortar y empalmar

En los organismos complejos, participan dos niveles de maquinaria molecular en el corte y empalme de los transcritos de pre-ARNm. La maquinaria basal, que se encuentra en todos los organismos cuyo genoma contiene intrones, se ha mantenido (o conservado) en el transcurso de la evolución: desde las levaduras hasta los humanos. Consta de cinco moléculas de ARN nuclear pequeño (ARNnp): U1, U2, U4, U5 y U6. Estas moléculas se reúnen con unas 150 proteínas para formar el somite cirujano ("spliceosome"), el complejo que se encarga de reconocer los sitios donde empiezan y terminan los intrones, podarlos, expulsarlos del transcrito de pre-ARNm y ensamblar entonces los exones para formar el ARNm.

En cada intrón, cuatro secuencias de nucleótidos operan como señales que indican al somite cirujano dónde cortar: al principio del intrón, o punto de corte 5'; al final, o punto de corte 3'; en medio, o zona de ramificación, y el tramo polipirimidínico.

Un sistema de regulación aparte controla el proceso de corte y empalme dirigiendo la maquinaria basal hacia estos puntos de corte. Se han identificado más de 10 proteínas reguladoras diferentes (proteínas SR, de "splicing regulatory"). Su forma varía según el tejido o la fase de desarrollo de un mismo tejido. Las proteínas SR se unen a cortas secuencias de nucleótidos de los exones del transcrito pre-ARNm. Estos puntos de unión se denominan potenciadores exónicos del corte y empalme (ESE, de "exonic splicing enhancer"), porque la unión cabal de la proteína SR a un ESE atrae a los ARNnp de la maquinaria basal a los puntos de corte adyacentes a uno u otro extremo del exón. Pero una proteína SR puede unirse también a una secuencia exónica supresora del corte y empalme (ESS, de "exonic splicing supressor"); cuando eso ocurre, la maquinaria basal pierde su capacidad de uncirse a los extremos de ese exón, que, por tanto, se cercena del ARNm.

El salto de un solo exón puede acarrear graves consecuencias para un organismo. En la mosca del vinagre, por ejemplo, el corte y empalme alternativo regula la vía de determinación del sexo. Cuando se expresa el gen *Sex-lethal*, puede saltarse por encima un exón específico del macho, durante el proceso de maduración; ello conduce a la síntesis

LUCY READING-IKKANDA

de una proteína Sex-lethal específica de la hembra. Esta proteína se une entonces a cualquier transcrito pre-ARNm subsiguiente del mismo gen, perpetuando la omisión del exón específico del macho en todos los procesos de corte y empalme posteriores, y, por tanto, garantizando la síntesis de la proteína específica de la hembra. En cambio, si el exón específico del macho es ensamblado durante la primera ronda de edición, se obtiene un ARNm no funcional, que dirige las células de la mosca hacia la ruta específica del macho.

La omisión de exones constituye el tipo de corte y empalme alternativo habitual en mamíferos. Pero no el único. En plantas y formas de vida multicelulares inferiores predomina un mecanismo que causa la retención de intrones en el ARNm maduro. La retención de intrones constituye a buen seguro la primera versión de edición alternativa en desarrollarse. Incluso hoy, la maquinaria de corte y empalme de organismos unicelulares, como las levaduras, opera mediante el reconocimiento de intrones; muy diferente el sistema de proteínas SR de los organismos superiores, que identifica los exones para marcarle el camino a la maquinaria basal.

En el sistema unicelular, la maquinaria de corte y empalme sólo reconoce secuencias intrónicas de menos de 500 nucleótidos. Esta forma de operar resulta idónea para la levadura, pues tiene pocos intrones, que comprenden unos 270 nucleótidos en promedio. Pero a medida que avanzaba la expansión de los genomas en el transcurso de la evolución, los tramos intrónicos se multiplicaron y crecieron. Quizá la maquinaria celular de corte y empalme se vio forzada a cambiar de un sistema que reconocía cortas secuencias intrónicas dentro de los exones a otro que reconociera exones de tamaño reducido en medio de un mar de intrones. Por término medio, un gen humano codificador de proteína consta de 28.000 nucleótidos, con 8,8 exones separados por 7,8 intrones. Los exones son bastante cortos, de unos 120 nucleótidos, mientras que los intrones oscilan entre 100 y 100.000 nucleótidos.

El tamaño y la cantidad de los intrones humanos —contamos con la mayor tasa de intrones por gen de todos los organismos— plantea una cuestión de interés. Los intrones resultan caros de mantener. Gran parte de la energía que consumimos a diario se dedica al mantenimiento y reparación de los intrones en su forma de ADN, en transcribir el pre-ARNm, en podar los intrones e incluso en la destrucción de los intrones al final del proceso de corte y empalme. Además, la edición puede ocasionar errores costosos. Cada corte y engarce de pre-ARNm erróneos comporta un cambio en la secuencia codificadora de proteína del transcrito del gen y, a buen seguro, la síntesis de una proteína defectuosa.

La disautonomía familiar, una enfermedad hereditaria, se debe a una mutación de un nucleótido en el gen *IKBKAP*; la mutación provoca una edición alternativa del mismo en tejidos del sistema nervioso. El consiguiente descenso en los niveles de proteína IKBKAP funcional conlleva un desarrollo anormal del sistema nervioso. Aproximadamente la mitad de los pacientes con esta enfermedad mueren antes de cumplir los 30 años. Al menos el 15 por ciento de las mutaciones génicas responsables de enfermedades genéticas (y probablemente también de algunos cánceres) operan mediante la alteración del corte y empalme del pre-ARNm. ¿Por qué, pues, habrá preservado la evolución un sistema tan complicado, capaz de causar enfermedad? Quizá porque las ventajas superan a los inconvenientes.

Las ventajas de lo alternativo

Al generar más de un tipo de ARNm y, por tanto, más de una proteína por gen, el corte y empalme alternativo permite a los humanos fabricar más de 90.000 proteínas sin tener que mantener 90.000 genes. Por término medio, cada uno de nuestros genes genera unos tres ARNm mediante edición alternativa. Sin embargo, este argumento no justifica que necesitemos tal cantidad de intrones ni que ocupen la mayor parte del espacio génico: dejan para los exones sólo un 1 o un 2 por ciento de nuestro genoma.

Tras revelarse aparentemente vacío este paisaje genómico en 2001, se presentó otro misterio desconcertante en 2002 con la publicación de la secuencia genómica del ratón. El múrido tenía un número de genes parecido al del hombre. Aunque de un antepasado común nos separan unos 100 millones de años, la mayoría de los genes humanos y del ratón derivan de ese precedente. En su mayoría, estos genes comparten la misma disposición de intrones y exones; la secuencia de nucleótidos en sus exones ha persistido, en buena medida, a través de la evolución. Si tan pequeña resulta, pues, la diferencia entre los genomas de humanos y múridos, ¿qué nos hace tan distintos de los roedores?

Christopher J. Lee y Barman Modrek, de la Universidad de California en Los Angeles, acaban de demostrar que una cuarta parte de los exones editados de forma alternativa en ambos genomas son específicos de humanos o de ratón. Estos exones tienen el potencial de fabricar proteínas específicas de especie, que serían responsables de la diversificación entre las especies. A este respecto, existe un grupo de exones editados alternativamente únicos en primates (humanos, simios y monos) que podrían haber contribuido a la divergencia de los primates respecto de los otros mamíferos. El estudio del proceso que da origen a exón permite vislumbrar las ventajas de los intrones en general. Veamos por qué compensa mantenerlos.

Estos exones específicos de primates derivan de elementos genéticos móviles Alu, que pertenecen a los retrotransposones. En esa clase extensa de elementos se acogen secuencias cortas de ADN que generan copias de ellas mismas, para reinsertarlas luego en el genoma en posiciones aleatorias, como si de pequeños parásitos genómicos se tratara. Los retrotransposones se encuentran en casi todos los genomas. Al contribuir a la expansión genómica, han ejercido una profunda influencia sobre la evolución de los organismos multicelulares. Casi la mitad del genoma humano está formado por elementos transponibles, siendo los Alu los más abundantes.

Los elementos Alu constan de 300 nucleótidos; su secuencia se caracteriza por terminar en una “cola poli A”. Nuestro genoma contiene unos 1,4 millones de elementos Alu, que en buena parte continúan multiplicán-

dose e insertándose en nuevos sitios a razón de una nueva inserción por cada 100 o 200 nacimientos en humanos.

Durante bastante tiempo, los Alu fueron considerados basura genómica. Empezaron a ganarse el respeto cuando se descubrió que la inserción de Alu aumentaba la capacidad de un gen de producir proteínas. Cerca de un 5 por ciento de los exones editados de forma alternativa en el genoma humano contienen una secuencia Alu. Esos exones se habrían originado cuando un elemento Alu "saltó" a un intrón.

En principio, ello no habría tenido consecuencia negativa para el primate, dado que, en su mayoría, los intrones suelen podarse y descartarse en el transcurso de la edición. Sin embargo, a través de una posterior mutación, el Alu habría transformado el intrón "huésped" en una secuencia de información genética con significado, es decir, en un exón. Esta transformación ocurre si en la secuencia Alu se producen cambios que crean un nuevo punto de corte, 5' o 3', dentro del intrón, puesto que ello hace que el somite cirujano reconozca parte del intrón como un "exón". (Tales mutaciones suelen aparecer mientras procede la división celular, cuando al copiarse el genoma se introduce una "errata".)

Si el nuevo exón Alu se conserva sólo de forma alternativa, el organismo saca partido de las dos posibles ediciones del pre-ARNm. Cuando el ARNm incluye el exón Alu, las células producen una nueva proteína; pero cuando no lo incluye (porque se ha omitido), se recupera la función original del gen. Un Alu mutado sólo resulta problemático si se conserva de forma constitutiva —es decir, si se incluye en todos los ARNm producidos a partir del gen—, pues la ausencia de la proteína primigenia puede provocar el desarrollo de enfermedades genéticas. Hasta la fecha, tres enfermedades genéticas causadas por una mala colocación de secuencias Alu han sido identificadas: los síndromes de Alport y de Sly y la anomalía de OAT.

Con mi grupo he demostrado que lo único que se necesita para convertir algunos elementos intrónicos Alu silentes en auténticos exones es el cambio de una sola letra en su secuencia de ADN. El genoma humano contiene alrededor de 500.000 elementos Alu alojados en intrones; 25.000 de ellos podrían convertirse en nuevos exones sólo con sufrir esa mutación puntual. Por tanto, las secuencias Alu encierran el potencial de continuar enriqueciendo nuestro repertorio de genes codificadores de proteínas.



2. CHIMPANCES Y HUMANOS comparten el 99 por ciento de su genoma. En ese apartado se incluyen los elementos móviles Alu, exclusivos de los primates. Los Alu pueden haber dado lugar, mediante el corte y empalme alternativo, a nuevas proteínas que habrían impulsado la divergencia de los primates respecto de otros mamíferos y quizá también la divergencia entre los humanos y otros primates. La investigación ha puesto de manifiesto que los genes casi idénticos de humanos y chimpancés producen esencialmente las mismas proteínas en la mayoría de los tejidos, excepto en el cerebro. En este órgano ciertos genes humanos son más activos y otros generan proteínas distintas mediante el corte y empalme alternativo.

MIKEL ROBERTS *Corbis Sygma*

Aplicaciones terapéuticas

Más de 400 laboratorios y unos 3000 expertos en todo el mundo tratan de desentrañar las complejas reacciones que participan en el corte y empalme alternativo. Aunque la investigación se encuentra en sus comienzos, los primeros resultados insinúan nuevas vías para futuras aplicaciones terapéuticas: por ejemplo, las nuevas terapias génicas que se sirven del mecanismo de corte y empalme para tratar patologías congénitas o adquiridas.

Un posible enfoque terapéutico se basaría en dirigir una corta secuencia de nucleótidos de ARN artificial o ADN sintético (oligonucleótido antisentido) para que se uniera a blancos específicos del ADN o ARN del paciente. Estos oligonucleótidos se introducirían en la célula para enmascarar o un punto de corte específico o cualquier otra secuencia reguladora, desplazando así la actividad de corte y empalme a otro sitio.

Ryszard Kole, de la Universidad de Carolina del Norte en Chapel Hill, aplicó esta técnica en células precursoras de la sangre procedentes de pacientes con beta talasemia, un trastorno hereditario en el cual un punto de corte 5' aberrante da lugar a malformaciones en las moléculas de hemoglobina (las encargadas de transportar el oxígeno). Enmascarando la mutación, Kole consiguió que el punto de corte recuperase su ubicación normal y con ello restaurar la producción de hemoglobina funcional.

Posteriormente, Kole aplicó la misma técnica a cultivos de células cancerosas humanas. Ocultando un

punto de corte 5' del *Bcl-x*, transcrito génico regulador de la apoptosis, desplazó la actividad de corte y empalme de manera que se generase la forma Bcl-x(S) de ARNm, no la forma Bcl-x(L); ello redujo la síntesis de la proteína antiapoptótica por parte de las células cancerosas y fomentó la síntesis de proteína proapoptótica. En algunas células cancerosas, este cambio activa el programa apoptótico; en otras, potencia los efectos apoptóticos de sustancias quimioterapéuticas administradas junto con los oligonucleótidos.

Otra aplicación terapéutica de la edición alternativa fue presentada en 2003 por Adrian Krainer y Luca Cartegni, del Laboratorio Cold Spring Harbor en Long Island. Hallaron un modo de inducir la retención de un exón, cuyo destino normal hubiera sido el descarte. Crearon una molécula sintética que se programó para unirse a cualquier tramo de ARN según su secuencia; le ensamblaron luego la parte de una proteína SR que se une al ARN. Esta molécula quimérica podía luego vincularse a una secuencia determinada del pre-ARNm y atraer la maquinaria basal hacia el punto de corte adecuado. Krainer y Cartegni aplicaron este método a células humanas cultivadas para corregir defectos de corte y empalme en versiones mutadas del gen *BRCA1*, implicado en el cáncer de mama, así como del gen *SMN2*, que causa atrofia muscular espinal.

Un tercer método aprovecha la capacidad de los somites cirujanos para ensamblar dos moléculas de pre-ARNm diferentes, procedentes del mismo gen, para formar un ARNm compuesto. Este proceso, denominado trans-corte y empalme, es frecuente en vermes, pero raro en humanos. Permite podar con precisión una región mutada de un pre-ARNm responsable de una enfermedad y reemplazarla por una secuencia codificadora de proteína normal. En fecha reciente, John Englehardt, de la Universidad de Iowa, ha empleado esta técnica *in vitro* para corregir parcialmente el pre-ARNm de un gen que produce una proteína defectuosa en las células de las vías respiratorias de los pacientes con fibrosis cística.

Antes de descodificarse el genoma humano, eran muy pocos los que creían que organismos de la complejidad humana pudieran sobrevivir con sólo 25.000 genes. Tras completarse la secuencia, sin embargo, el corte y empalme alternativo ha emergido como el proceso fundamental que permite a un reducido acervo génico generar un vasto repertorio de proteínas, necesarias para constituir el cuerpo humano, al tiempo que orquesta de forma precisa la fabricación de las mismas en los distintos tejidos y en las diferentes etapas. Además, el corte y empalme explica de qué modo pudo originarse, a partir de genomas similares, la rica diversidad que existe entre humanos, múridos y, presumiblemente, todos los mamíferos.

La evolución opera al presentarles a los organismos nuevas posibilidades y la selección de los individuos más idóneos. Por tanto, las proteínas creadas mediante el ensamblaje de nuevos exones derivados de Alu ayudaron, sin duda, a convertir la especie humana en lo que es hoy. La investigación acerca del corte y empalme alternativo promete reforzar nuestra calidad de vida.

El autor

Gil Ast es catedrático del departamento de genética humana y medicina molecular en la facultad de medicina de la Universidad de Tel Aviv. Centra su investigación en el mecanismo molecular de corte y empalme del pre-ARNm.

Bibliografía complementaria

ALTERNATIVE SPLICING: INCREASING DIVERSITY IN THE PROTEOMIC WORLD. B. R. Graveley en *Trends in Genetics*, vol. 17, n.º 2, págs. 100–107; febrero, 2001.

SPLICING REGULATION AS A POTENTIAL GENETIC MODIFIER. M. Nissim-Rafinia y B. Kerem en *Trends in Genetics*, vol. 18, n.º 3, págs. 123–127; marzo, 2002.

LOS INTRONES. John S. Mattick en *Investigación y Ciencia*, págs. 26-33; diciembre, 2004.

HOW DID ALTERNATIVE SPLICING EVOLVE? Gil Ast en *Nature Reviews Genetics*, vol. 5, págs. 773–782; octubre, 2004.

Transmisión de secretos nucleares

Durante la segunda guerra mundial

Cuando la agencia TASS comunicó oficialmente el 25 de septiembre de 1949 que la Unión Soviética disponía ya de la bomba atómica, los Estados Unidos se quedaron sorprendidos, pues aunque sabían que los soviéticos tendrían tarde o temprano armas nucleares, no esperaban que fuera tan pronto. Cuando el 27 de enero de 1950 el científico alemán Klaus Fuchs, miembro del Proyecto Manhattan, declaró ante Scotland Yard, la situación quedó bastante clara.

El 5 de septiembre de 1945 el diplomático Igor Guzenko, máximo responsable de la información cifrada en la embajada soviética de Canadá, pidió asilo político en este país a cambio de unos importantísimos documentos que revelaban la existencia durante la guerra de una amplia red de espionaje soviético en Canadá y en los Estados Unidos. Esa información llevaría con el tiempo a la detención de los científicos nucleares Klaus Fuchs y Alan Nunn May.

La convicción de un espía

Klaus Fuchs, activo militante comunista huido de la Alemania nacionalsocialista, participó en el muy secreto Proyecto Manhattan. Transfirió información nuclear a los servicios de inteligencia de la Unión Soviética entre 1942 y 1949, año en que fue detenido.

Se llegó a decir de él que fue el espía más importante del siglo XX. Emprendió ese camino movido por la profunda convicción moral e ideológica que le ligaba a la Unión Soviética. A diferencia de otros agentes, nunca aceptó dinero a cambio.

El Directorio de Aleaciones para Tuberías, en realidad el proyecto nuclear británico, empezó sus trabajos en 1941. Klaus Fuchs formó allí parte del equipo de investigación dirigido por Rudolph Peierls, científico alemán judío también refugiado en Inglaterra. En una entrevista realizada en Washington en 1989, pregunté a Sir Peierls sobre Fuchs. Me contó que fue un buen científico, competente y rápido, pero muy introvertido.

En 1941, cuando el III Reich invadió la URSS, Fuchs entró en contacto con el agente de inteligencia Simon D. Kremer, que se encargaría de hacer llegar sus comunicaciones verbales y notas escritas por medio de la embajada soviética en Londres.

El 19 de agosto de 1943 se firmó en Quebec el Acuerdo de Cooperación para la Energía Atómica entre los Estados Unidos y el Reino Unido, por el que ambos países se comprometían a intercambiarse información científico-técnica. Debido a este acuerdo secreto, el equipo científico inglés se insertó en el Proyecto Manhattan. Fuchs se encuadró en la División de Física Teórica, una de las más importantes, dirigida por Hans Bethe, que luego sería premio Nobel. Tuve ocasión de hablar personalmente con el profesor Bethe en su despacho del Instituto Newmann de Estudios Nucleares en la Universidad de Cornell. Me dijo que Fuchs fue un físico muy capaz y trabajador en lo que atañía a su capacidad científica, pero nadie llegó a intimar con él.

En 1946 Klaus Fuchs regresó a Inglaterra, donde se le nombró inmediatamente director de la División de Física Teórica del Centro Británico de Investigaciones Nucleares de Harwell. Ese mismo año se produjo el arresto del profesor Israel Halperin, cuyo nombre constaba en los archivos que el diplomático Guzenko entregó al gobierno canadiense, y ello condujo a la detención de Fuchs en 1949 por parte de William James Skardon, un brillante agente de contraespionaje de Scotland Yard, que paulatinamente había ido acercándose a Fuchs hasta ganarse su confianza personal. Esta labor culminó el 27 de enero de 1950 cuando Klaus Fuchs firmó la confesión oficial.

Cuando lo detuvieron, Fuchs declaró que había pasado informes completos a los soviéticos sobre las investigaciones llevadas a cabo tanto en Inglaterra como en los Estados Unidos. En estos informes se incluían datos sobre la alta tasa de neutrones producidos en la fisión espontánea del ^{240}Pu. Esa característica impedía aplicar el método del proyectil, empleado en la bomba de uranio, en la bomba de plutonio. Esta requería el método de implosión. También dio datos sobre la masa crítica del ^{240}Pu, unos 5 kg, inferior a la de ^{235}U. Cuando el proyecto de la bomba de plutonio quedó concluido, Fuchs pasó información de sus dimensiones, pesos y características. Fue acusado de violación del Acta de Secretos Oficiales y no de traición, ya que no había pasado información a un país enemigo sino a un país aliado de guerra, que era lo que ha-



Fotografía del expediente de Klaus Fuchs en Los Alamos (1943). (Cortesía de Los Alamos National Laboratory.)

bía sido la Unión Soviética. Se le condenó a 14 años de prisión, pero fue liberado 9 años más tarde por buena conducta. El 23 de junio de 1959 salió de la cárcel de Wakefield con destino a Berlín Oriental. En Alemania del Este siguió con sus investigaciones como subdirector del Instituto de Física Nuclear en Rossendorf, próximo a Dresde. Klaus Fuchs, uno de los espías más importantes del siglo XX, falleció en Berlín el 28 de enero de 1988.



Edificio de la Lubianka en Moscú que albergaba el Comisariado del Pueblo para Asuntos Internos (NKVD). Desde 1917 constituye la sede permanente de los órganos de seguridad del Estado y de sus servicios secretos.

Otras fuentes de información

Alan Nunn May trabajó para el Proyecto Manhattan en Canadá. Formaba parte de Aleaciones para Tuberías desde 1942. Al llegar a Canadá como miembro de la misión científica británica dirigida por el profesor John Cockcroft, fue reclutado como agente al servicio del NKVD —el organismo de inteligencia soviético precursor del KGB— por el coronel Nicolai Zabotin, agregado militar en la embajada de Ottawa. Nunn May informó de las actividades de los laboratorios de Montreal y Oak Ridge —éste en Estados Unidos—, que albergaban las plantas de separación isotópica del uranio.

Aunque se transfirió información nuclear durante la guerra por varios conductos, podríamos decir que, por lo sabido hasta ahora, Fuchs fue sin duda el espía más importante cualitativa y cuantitativamente hablando. No obstante, según el teniente coronel Anatoly Iatskov, antiguo jefe del espionaje soviético en Nueva York, dentro del Proyecto Manhattan hubo un espía norteamericano, *Perseus*, más importante que Fuchs. Su identidad permanece bajo el secreto del KGB. Por otra parte, el coronel Vladimir Barkovski, que en la década de 1940 fue uno de los principales agentes soviéticos que recopilaron abundantes secretos acerca de la investigación nuclear que se llevaba a cabo en Occidente, sostenía que era imposible que una cantidad tan grande de información proviniese sólo de Klaus Fuchs. Según Barkovski, había al menos diez expertos británicos que suministraban a la Unión Soviética información sobre la bomba atómica y cuyas identidades permanecen en secreto.

En 1996 la Agencia de Seguridad Nacional norteamericana (NSA) hizo público un comunicado en el que identificaba a Theodore Hall, profesor en Cambridge fallecido en 1999, como otro de los espías que había trabajado para Moscú. Según la agencia de inteligencia norteamericana, Hall, apodado *Mlad*, fue uno de los agentes que aportaron a los soviéticos información, entre otras cosas, de la primera prueba de una bomba atómica, en Alamogordo (Nuevo México) el 16 de julio de 1945. Se la hizo llegar por mediación de los agentes británicos Morris Cohen (conocido también con el alias de *Peter Kroger*) y su esposa Lona. Su declaración queda ampliamente recogida en el libro *Bombshell: The Secret Story of America's Unknown Atomic Spy Conspiracy*.

Tras la detención de Klaus Fuchs, el FBI se puso a trabajar febrilmente buscando los contactos que Fuchs hubiera podido tener en los Estados Unidos. A partir de aquí comenzó a tramarse una historia que quedará para siempre sumida en contradicciones e interrogantes y que concluiría con la ejecución en la silla eléctrica del matrimonio Rosenberg el 19 de junio de 1953. Se los acusó de espionaje y de pasar a la Unión Soviética información crucial del Proyecto Manhattan que les había sido proporcionada por el hermano de Ethel Rosenberg, David Greenglass. Sin embargo, la realidad era que la información que transmitieron tenía escaso valor técnico.

No obstante, la trama de espionaje a favor de la Unión Soviética



Georgi N. Flerov, uno de los científicos más importantes del proyecto nuclear soviético. En 1941 descubrió junto con su compañero Konstantin a. Petrjak la fisión espontánea del uranio. Al fondo de la foto puede verse un retrato de Igor V. Kurchatov.

por parte de científicos occidentales sigue encerrando grandes interrogantes. Aún se escribe sobre ello. El libro *Special Tasks* (1994), de Pavel Sudoplatov, antiguo director de espionaje y de operaciones especiales del servicio de inteligencia soviético, dedica el capítulo siete al espionaje nuclear durante la guerra. Ha sido muy criticado por contener información desorientadora, y difícil de probar hasta la fecha, que involucra a Robert Oppenheimer, director científico del Proyecto Manhattan, y a Enrico Fermi. En la actualidad siguen surgiendo más acusaciones, públicamente rechazadas por los propios implicados o sus familiares.

La destrucción mutua asegurada

El primer proyecto nuclear soviético fue dirigido por el físico nuclear Igor V. Kurchatov, un científico eminente. En su equipo trabajaron científicos de primera clase: Zeldovich, Jaritón, Flerov, Petrjak, Alijanov, Pervugin, etc. La información que recibieron de los servicios secretos les fue útil, pero sólo sirvió para adelantar un par de años el resultado que habrían obtenido por sus propios medios. Kurchatov la utilizó con cautela, pues pensaba que podía haber sido amañada.

La explosión de la bomba atómica de la URSS tuvo como consecuencia inmediata la fabricación de la bomba de hidrógeno por parte de ambas potencias, soviética y norteamericana. Se recrudeció la guerra fría y nació la política de la “destrucción mutua asegurada”.

NATIVIDAD CARPINTERO SANTAMARÍA
Instituto de Fusión Nuclear
Universidad Politécnica de Madrid

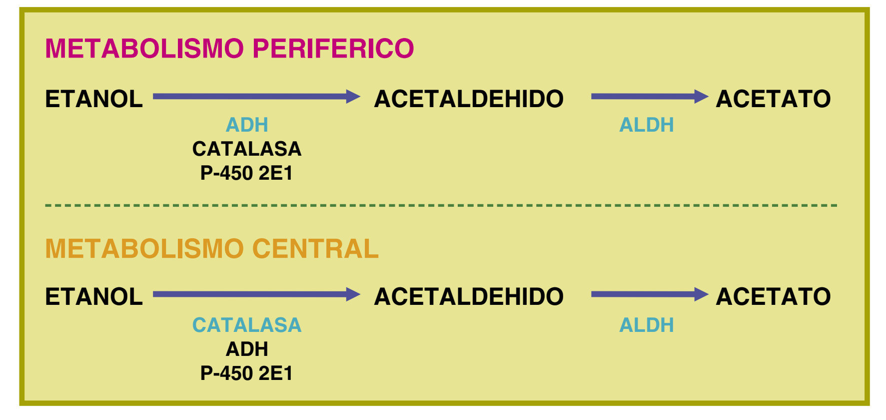
Metabolismo cerebral del alcohol

Función de la enzima aldehído deshidrogenasa

Una de las cuestiones debatidas en la bioquímica del etanol, con sus implicaciones adictivas consiguientes, se centra en el metabolismo cerebral de la molécula. Para encuadrar el tema, resulta obligado arrancar de la biosíntesis del etanol.

En el proceso de síntesis del alcohol se hallan implicados diversos órganos (riñones, hígado y estómago, principalmente). Corresponde a la alcohol deshidrogenasa promover la transformación de etanol en acetaldehído. Una pequeña cantidad de etanol, el 10 % o menos, puede biosintetizarse por rutas alternativas, como la del citocromo P-450 2E1 o la catalasa. En una fase terminal de la ruta de biosíntesis, catalizada por la aldehído deshidrogenasa, se obtiene acetato a partir del acetaldehído.

BIOSINTESIS DEL ALCOHOL



1. Las enzimas que participan en la transformación del etanol tienen una eficacia diferencial, dependiente de la estructura donde se localicen. Lo que llamamos nivel periférico comprendería el hígado, los riñones y otros órganos, mientras que el nivel central remite al encéfalo en su conjunto. En la figura se reflejan con distintos colores las enzimas que participan en el metabolismo del etanol. En la periferia del organismo, las enzimas distintas de la alcohol deshidrogenasa (ADH) y la aldehído deshidrogenasa (ALDH) ejercen un papel secundario. En el cerebro, la catalasa “suplanta” las funciones de la alcohol deshidrogenasa, al tiempo que el sistema enzimático de aldehído deshidrogenasa trabaja a pleno rendimiento.

Desde un comienzo el acetaldehído fue considerado una sustancia con efectos aversivos. Acumulado en la periferia del organismo, se aprovecharon sus propiedades repulsivas para prevenir el consumo de alcohol en personas adictas. Pero la investigación pone ahora sobre el tapete una vertiente desconocida: su capacidad de producir efectos reforzantes.

Avalan la capacidad reforzante del acetaldehído ciertos ensayos con roedores. Los animales, por un lado, se autoadministran la sustancia; por otro, el acetaldehído produce preferencia de lugar tras administraciones intraperitoneales o intracerebroventriculares. Téngase en cuenta que ambos protocolos experimentales (autoadministración y preferencia de lugar) se emplean para la evaluación de drogas reforzantes y adictivas, como la cocaína. Cabría, pues, que el acetaldehído mediara algunas acciones psicofarmacológicas que se atribuyen al alcohol.

Pero esa sugerencia encontraba una objeción: la presencia de aldehído deshidrogenasa en el sistema microvascular cerebral, enzima que se interpone e impide que el acetaldehído periférico atraviese la barrera hematoencefálica. En ese caso, carece de sentido hablar de los efectos farmacológicos que pudieran lograr las pequeñas cantidades de acetaldehído que llegaran hasta el cerebro.

Pero, ¿y si se metabolizara el etanol directamente en el cerebro? ¿Podrían ser la catalasa y la aldehído deshidrogenasa encefálicas las enzimas encargadas del proceso transformador?

En el cerebro las actividades enzimáticas se invierten. La principal ruta de metabolismo de etanol que, en el hígado, está constituida por la alcohol deshidrogenasa, presenta en el cerebro una actividad casi nula; y ello, pese a encontrarse la enzima en varias estructuras cerebrales. Hallamos también en el encéfalo los mecanismos necesarios para el metabolismo oxidativo del etanol; en concreto, la catalasa y la aldehído deshidrogenasa.

En resumen, nada impedía que el acetaldehído se sintetizara directamente en el tejido cerebral.

Síntesis de acetaldehído

Para abordar la síntesis cerebral de acetaldehído se procedió a inhibir la acción de la catalasa por 3-amino-1,2,4-triazol. Cuando administramos este inhibidor no competitivo de la enzima, hay que esperar hasta que la catalasa se sintetice de nuevo para que pueda ejercer su acción enzimática. A través de la aplicación de este procedimiento se han obtenido cambios en la actividad locomotora, la letalidad y otras conductas inducidas por etanol.

Dentro del dominio de los sistemas enzimáticos relacionados con el metabolismo del etanol, la autora se ha centrado en la aldehído deshidrogenasa. Nos propusimos determinar si una mayor cantidad de acetaldehído encefálico permitiría la inducción de conductas atribuidas al etanol. En concreto, nos ceñimos a la actividad locomotora inducida por etanol, si bien con la perspectiva última de hallar un nuevo respaldo a la hipótesis del metabolismo cerebral del etanol.

Nos apoyamos en el dietilditiocarbamato (DDTC), un metabolito *in vivo* del disulfiram (fármaco habitual en el tratamiento del alcoholismo). El DDTC constituye un inhibidor de la actividad hepática y cerebral de la aldehído deshidrogenasa mitocondrial. Este isozima se caracteriza por presentar una constante de afinidad por el alcohol muy pequeña; basta una exigua concentración de alcohol para que la aldehído deshidrogenasa comience su acción metabólica y altere los niveles periféricos y centrales de acetaldehído.

A menudo se recurre también a la cianamida, un inhibidor conjunto de la aldehído deshidrogenasa y la catalasa, y al 3-amino-1,2,4-triazol, un inhibidor de la catalasa, para establecer qué función desempeñan ambas enzimas en la mediación de algunos efectos psicofarmacológicos del etanol.

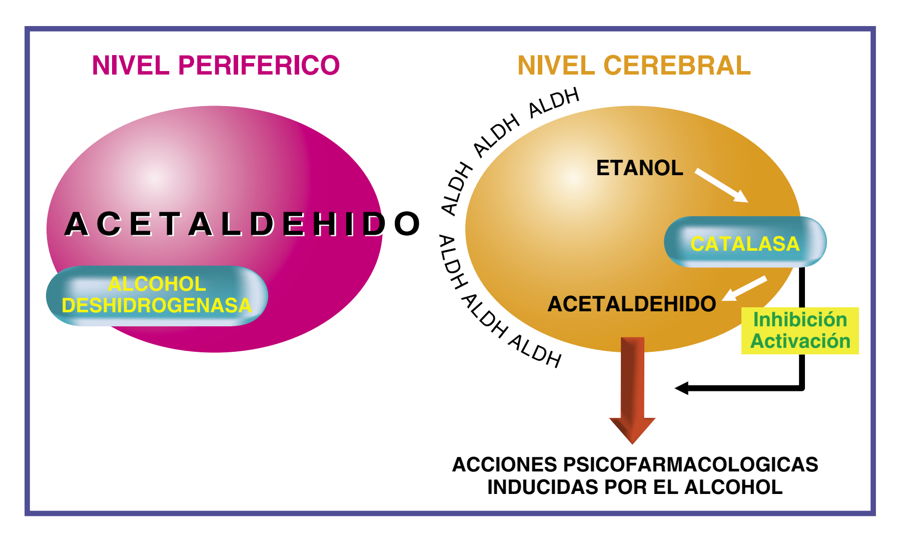
Con todo, podemos acotar mejor la acción de la aldehído deshidrogenasa en el metabolismo del etanol si empleamos DDTC, un inhibidor específico de su acción. Hemos ensayado también la administración concurrente de DDTC y 4-metilpirazol (4-MP); este último es un inhibidor de la alcohol deshidrogenasa.

El 4-MP bloquea la eliminación de etanol, al impedir la acumulación de acetaldehído durante el metabolismo del alcohol. De hecho, se ha venido aplicando ya el 4-MP para reducir la acumulación de acetaldehído periférico después de la administración de disulfiram. Con el bloqueo de la oxidación de etanol se busca inhibir la síntesis de acetaldehído.

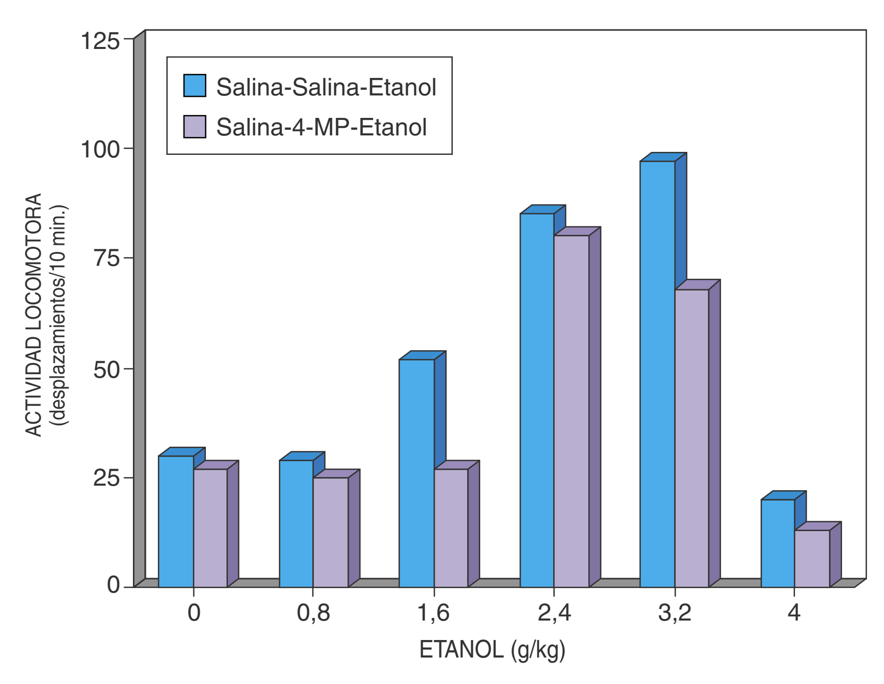
Entre sus efectos, el 4-MP disminuye la acumulación periférica de acetaldehído y revierte la taquicardia, el enrojecimiento facial y la reacción electrocardiográfica producida por la administración de inhibidores de aldehído deshidrogenasa (disulfiram, cianamida o nitrefazole).

Dimos un paso más: convertimos la capacidad del 4-MP para reducir los niveles de acetaldehído periféricos después de la administración de etanol en una herramienta para analizar el metabolismo cerebral del etanol. Los resultados de la interacción conjunta entre DDTC y 4-MP fueron determinantes. Hasta entonces nadie había optado por una manipulación farmacológica que mostrara los efectos psicobiológicos de ese tratamiento concurrente.

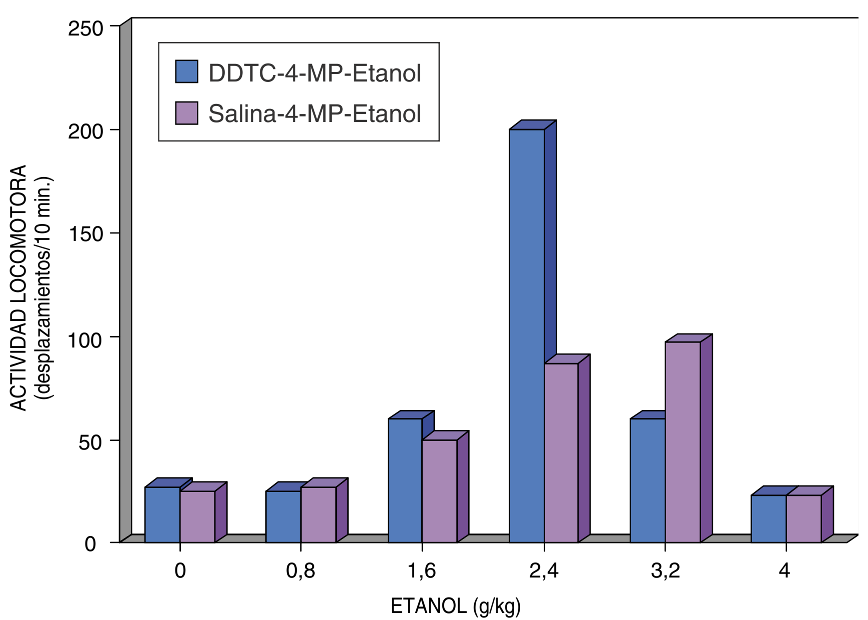
2. El acetaldehído, el primer metabolito de la síntesis del alcohol, ejerce no sólo efectos aversivos, sino también reforzantes. De los primeros se aprovechan los fármacos interdictores en el tratamiento del alcoholismo. A los efectos reforzantes se debe que el alcohol se consuma de forma compulsiva y los sujetos desarrollen dependencia.



3. En el metabolismo periférico del alcohol, la enzima más eficaz es la alcohol deshidrogenasa, que transforma en acetaldehído un 90 % del alcohol que ingerimos. Derivado de la biosíntesis periférica del alcohol, el acetaldehído encuentra dificultades para cruzar la barrera hemoencefálica. Se lo impide la aldehído deshidrogenasa (ALDH), enzima cuya acción transforma el aldehído en acetato. Cabía, pues, sospechar la síntesis cerebral directa de acetaldehído a través del metabolismo encefálico del alcohol. Y, dado que en el tejido cerebral existen enzimas, como la catalasa y la ALDH, que intervienen en la síntesis del alcohol, y cuya activación o inhibición conlleva cambios en las acciones inducidas por esta droga, el acetaldehído podría mediar algunos de los efectos psicofarmacológicos inducidos por el alcohol.



4. Se muestra la acción del 4-metilpirazol (4-MP) sobre la actividad locomotora inducida por etanol. Los animales fueron pretratados con inyecciones intraperitoneales de solución salina o 4-MP, cinco minutos antes del tratamiento con diferentes dosis de etanol.



5. Se muestra el efecto de la administración concurrente de dietilditiocarbamato (DDTC) y 4-metilpirazol (4-MP) o salina sobre la actividad locomotora inducida por etanol. Los animales fueron pretratados con inyecciones intraperitoneales de DDTC y 4-MP o solución salina y 4-MP, ocho horas antes del tratamiento con diferentes dosis de etanol.

Los datos

Se administró a ratones adultos DDTC y 4-MP. Si se administraba sólo 4-MP, no repercutía en la actividad locomotora espontánea ni en la actividad inducida por etanol. Si se ofrecía una solución salina y 4-MP, los animales mostraban una conducta locomotora similar a los ratones pretratados sólo con la solución salina. El mismo fenómeno que se producía con la administración de 4-MP solo, se repetía con la administración de DDTC solo.

Pero, ¿qué acontecía si se les inyectaba primero el pretratamiento conjunto DDTC y 4-MP y, luego, alcohol? Asistíamos a un incremento significativo de actividad locomotora inducida por etanol. No se registró, sin embargo, modificación de la actividad locomotora espontánea en ninguno de los grupos controles. Quedaba claro que el incremento en la actividad locomotora no podía atribuirse al tratamiento con 4-MP solo, ni debía asociarse a un incremento en los niveles periféricos del etanol.

En otro orden, el efecto inhibitorio en la actividad de la alcohol deshidrogenasa adquiere interés en el caso exclusivo del metabolismo periférico del etanol. La acción de esa enzima es prácticamente nula en el metabolismo cerebral del alcohol.

En breve, del conjunto de los resultados anteriores se desprendía el valor de 4-MP como herramienta farmacológica combinada con DDTC para demostrar la importancia del acetaldehído y sus enzimas biosintetizadoras en la mediación de la actividad locomotora inducida por etanol.

Aldehído deshidrogenasa

La alcohol deshidrogenasa cerebral desempeña un papel menor en el metabolismo central del etanol. En cambio, la inhibición de aldehído deshidrogenasa causada por el DDTC modifica la actividad de ésta en la periferia del organismo y en el cerebro. La acción del DDTC sobre la aldehído deshidrogenasa da lugar a niveles de acetaldehído periférico y cerebral mayores de lo normal.

Sin embargo, el efecto de 4-MP sobre la alcohol deshidrogenasa podría reducir los niveles de acetaldehído periférico, sin que con ello ejerciera un efecto significativo sobre los niveles de acetaldehído encefálico.

El aumento de concentración de acetaldehído cerebral se debe a la inhibición de una isozima, la aldehído deshidrogenasa mitocondrial ($ALDH_2$). Responsable de ese bloqueo es la acción inhibitoria de DDTC, junto con la acción concurrente de 4-MP. Al rebajarse los niveles periféricos de acetaldehído, podría favorecerse la expresión conductual de las acciones del etanol. Ello nos lleva a un incremento signi-

ficativo en el efecto del etanol sobre la actividad locomotora.

Los datos recogidos podrían poner de manifiesto estos efectos, ya que la administración concurrente de DDTC y 4-MP anula el efecto inhibitorio del DDTC sobre la actividad locomotora inducida por etanol. El incremento resultante en la acción del etanol abona la idea de que la aldehído deshidrogenasa cerebral desempeña, sin duda, un importante papel en la mediación de la actividad locomotora inducida por etanol. La enzima regularía los niveles de acetaldehído cerebral y promovería, a su vez, la viabilidad de un metabolismo cerebral del alcohol.

Mª Dolores Escarabajal Arrieta
Area de Psicobiología,
Universidad de Jaén

La naturaleza del ADN

Determina una correcta segregación de los cromosomas en la mitosis

Mientras que los acontecimientos que caracterizan a la mitosis en sus aspectos morfológicos se conocen desde hace mucho tiempo, dada la espectacularidad del comportamiento cromosómico y su fácil observación, no podemos decir lo mismo de los mecanismos moleculares que, en una actuación secuencial de una serie de genes de control que aún son objeto de investigación, van determinando de manera exquisita el progreso a través del ciclo celular en general y de la mitosis en particular. Una de las cuestiones sin duda más oscuras es cómo tiene lugar una correcta separación (segregación) de las dos dobles cadenas de ADN que resultan de la replicación durante el período S de la interfase de modo que en la mitosis se encuentre cada una de ellas en una de las cromátidas del cromosoma metafásico. Si tenemos en cuenta la longitud del ADN advertiremos la dificultad que entraña este proceso que ha de ser por una parte rápido y al tiempo admirablemente exacto, ya que los errores se pagan caros. La cuestión es: ¿qué enzima o enzimas se encargan de esta delicada misión de manera que el proceso esté en lo posible libre de errores?

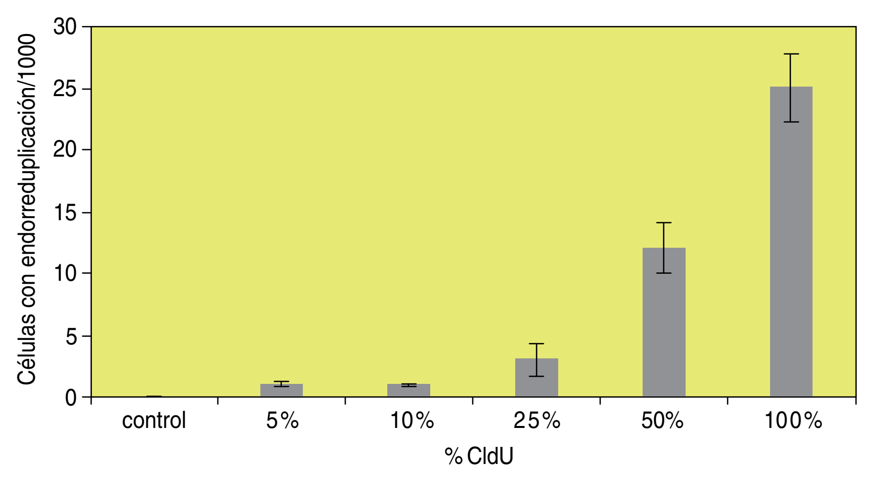


1. Diplocromosomas (4 cromátidas en lugar de 2) en una metafase tras producirse endorreduplicación por fallo en la segregación cromosómica como resultado de sustitución de timidina por CldU en el ADN de un fibroblasto de ovario de hámster chino en cultivo. La tinción diferencial en tonos claro y oscuro es indicativa de un alto grado de incorporación del nucleósido halogenado en el ADN.

Las topoisomerasas de ADN (topos) son enzimas nucleares muy conservadas evolutivamente que catalizan una variedad de cambios topológicos del ADN fundamentales durante la replicación, transcripción y recombinación. Estas enzimas cortan una de las dos cadenas de la molécula (topo I) o bien ambas (topo II) y tras permitir el paso de la otra cadena o cadenas intactas por la mella creada sellan ellas mismas la rotura. Se logra así relajar el superenrollamiento del ADN que se genera en los procesos arriba mencionados. Una diferencia muy interesante entre ambas topoisomerasas es que exclusivamente la de tipo II es capaz de desencadenar dobles cadenas de ADN, del tipo ya replicado al que nos referimos al principio. Dado que en mutantes de levadura con topo II defectuosa no se lleva a cabo la segregación cromosómica, todo apunta a que dicha enzima es la principal responsable en dicho proceso.

El ciclo catalítico de la enzima topo II para llevar a cabo la topoisomerización del ADN empieza mediante la unión de la misma a su sustrato de ADN bicatenario, proceso que depende de la secuencia de nucleótidos del ADN, así como directamente del mayor o menor grado de plegamiento de la molécula. En cuanto a lo primero, si bien no se ha encontrado hasta ahora una alta especificidad, la topo II corta ambas cadenas del ADN en secuencias preferentes dentro de sus sitios de reconocimiento/unión a la molécula. En cualquier caso, no obstante lo anterior, puede decirse que las reglas que determinan la especificidad de la topo II por determinadas secuencias nucleotídicas en el ADN se desconocen aún en buena medida.

Volviendo a la segregación cromosómica y al posible papel de la topo II, una de las manifestaciones más espectaculares de un fallo en la segregación es la observación de diplocromosomas, los cuales presentan en la metafase cuatro cromátidas en lugar de dos (*véase la figura 1*). Ello se debe al fenómeno de la endorreduplicación, que consiste en dos rondas sucesivas de replicación del ADN sin mitosis entre ellas al no poder segregarse correctamente las



2. Paralelismo entre el porcentaje de sustitución de la timidina en el ADN por un nucleósido halogenado (en este caso CldU) y frecuencia de endorreduplicación observada.

cromátidas por problemas en la separación de los dúplices de ADN tras completarse la replicación.

Se ha publicado que mediante fármacos que interfieren con la topo II se induce la endorreduplicación, lo cual sin duda constituye una prueba adicional de la responsabilidad de la enzima en la segregación cromosómica. Mediante un abordaje alternativo, es decir, sin emplear ningún agente que actúe sobre la topo II, en el grupo de investigación "Cultivo celular y radiobiología" de la Universidad de Sevilla hemos obtenido resultados que apuntan claramente al papel determinante de la naturaleza del ADN en la segregación cromosómica. La idea de partida era que si realmente existe una especificidad de la topo II por determinadas secuencias de nucleótidos en el ADN para su unión y posterior corte, una alteración importante en dichas secuencias podría resultar en una notable falta de actividad de la enzima; ello se traduciría en fallos en la segregación detectables mediante un incremento en la endorreduplicación. En este sentido, nuestra experiencia en el dominio de las condiciones experimentales para el control del grado de sustitución en el ADN de la timidina por otros nucleósidos, en este caso halogenados, ha sido un factor decisivo. La elección de dichos nucleósidos halogenados para alterar la naturaleza del ADN tiene la ventaja de que se puede detectar su incorporación en los cromosomas, al producirse alteraciones en la cromatina que resultan en su tinción diferencial.

Elegimos, pues, tres nucleósidos halogenados para su incorporación en el ADN en sustitución de la timidina y observamos el posible efecto en la segregación cromosómica. En todos los casos aparecieron metafases con diplocromosomas. Debíase ese resultado a una mala segregación, siendo la clorodesoxiuridina (CldU) el halogenado más eficaz (*véase la figura 2*), frente a la yododesoxiuridina (IdU) y bromodesoxiuridina (BrdU). Además, la frecuencia de metafases que presentan diplocromosomas muestra un buen paralelismo con el porcentaje preestablecido de sustitución de la timidina por el halogenado (*véase la figura 2*). Semejante paralelismo ofrece un apoyo adicional al conjunto de resultados en lo que se refiere a la necesidad de que el ADN se encuentre en su estado natural para que la segregación funcione normalmente.

Una observación muy interesante ha sido la siguiente: cuando el análogo halogenado se incorpora sólo en el ADN naciente no se aprecia un incremento en la endorreduplicación, lo cual parece indicar que es la naturaleza del ADN parental lo que determinaría un correcto reconocimiento por parte de la topo II para su insustituible función en la segregación cromosómica. No obstante, dada la afinidad de la topo II por el ADN superenrollado frente al relajado, no puede descartarse que nuestras observaciones se deban a diferencias en la condensación de la cromatina ocasionadas por la sustitución de la timidina por nucleósidos halogenados en el ADN.

FELIPE CORTÉS BENAVIDES
Depto. Biología Celular,
Facultad de Biología
Universidad de Sevilla

GLOSARIO

Nucleósido: en el ADN, el compuesto (o unidad) formado por la unión de cualquiera de las cuatro bases que lo componen (púrica o pirimidínica) con el azúcar desoxirribosa.

Secuencia nucleotídica: secuencia de las bases del ADN enlazadas con desoxirribosa y fosfato (nucleótidos).

Mitosis: fase del ciclo celular en que se divide la célula y los nuevos cromosomas se reparten entre las dos células resultantes.

Interfase: período entre dos mitosis.

Período S de la interfase: parte de la interfase en que se copia el ADN (la S se refiere a esa síntesis de ADN) y duplica cada cromosoma.

Cromátidas: hebras cromosómicas —procedentes de la duplicación, durante el período S de la interfase, de un solo cromosoma— que se condensan durante la primera parte de la mitosis o profase, se unen en un estrechamiento —el centrómero— durante la segunda o **metafase**, y se desagregan durante la última o anafase.

Superenrollamiento del ADN: enrollamiento del eje en torno al que se enrollan las hebras de la molécula de ADN.

De cerca

Francesc Peters

Temporal dentro del agua



1. Temporal de levante en una zona costera de Barcelona (playa de la Barceloneta) en noviembre de 2001.

Tememos a los temporales marinos porque dificultan la navegación, nos impiden disfrutar de la playa y pueden causar daños en la costa. Pero no solemos pensar en las consecuencias que comportan para los organismos que medran en este fluido acuoso de movimiento incesante. En el mar, el agua fluye turbulenta; su flujo se caracteriza por la permanente generación y destrucción de remolinos de diferente tamaño. La intensidad de la turbulencia viene determinada por las entradas de energía en el sistema, como el viento que incide sobre la superficie del mar. Así pues, lo mismo que las hojas que del suelo el viento levanta en remolinos durante un vendaval, los organismos marinos dotados de escasa o nula capacidad natatoria (plancton, sobre todo) se ven zarandeados y arrastrados por remolinos de agua.

En función de la intensidad de zarandeo, el efecto sobre el plancton puede ser beneficioso o perjudicial. Una turbulencia moderada aumenta el flujo de nutrientes hacia el fitoplancton e incrementa las tasas de encuentro entre depredadores planctónicos y sus presas, contribuyendo así al aumento de la producción de estas poblaciones. Con una turbulencia muy intensa, se resienten los procesos de captura y manipulación de presas, pueden causarse cambios fisiológicos, alteraciones morfológicas, interrupción del ciclo de división celular y modificaciones de los patrones de bioluminiscencia. Los efectos de la turbulencia sobre el plancton se estudian sobre todo en el laboratorio. Las turbulencias se pueden simular mediante una rejilla que oscila verticalmente en un contenedor; variando la amplitud y la frecuencia de dicha oscilación se generan turbulencias de distinta intensidad.

La supervivencia del plancton marino —que se caracteriza por un tiempo de vida de entre un día y unas pocas semanas— viene condicionada por la adaptación a niveles basales de zarandeo constante y al tiovivo ocasional de los episodios de turbulencia. La próxima vez que presenciemos un temporal marino, podemos preguntarnos cómo lo estará capeando el plancton.

PEP GASOL



2. Temporal en el Paso de Drake (entre el continente sudamericano y la península Antártica) visto desde el buque de investigación oceanográfica BIO Hespérides. Este es uno de los mares más agitados de mundo; sufre temporales de manera casi constante.

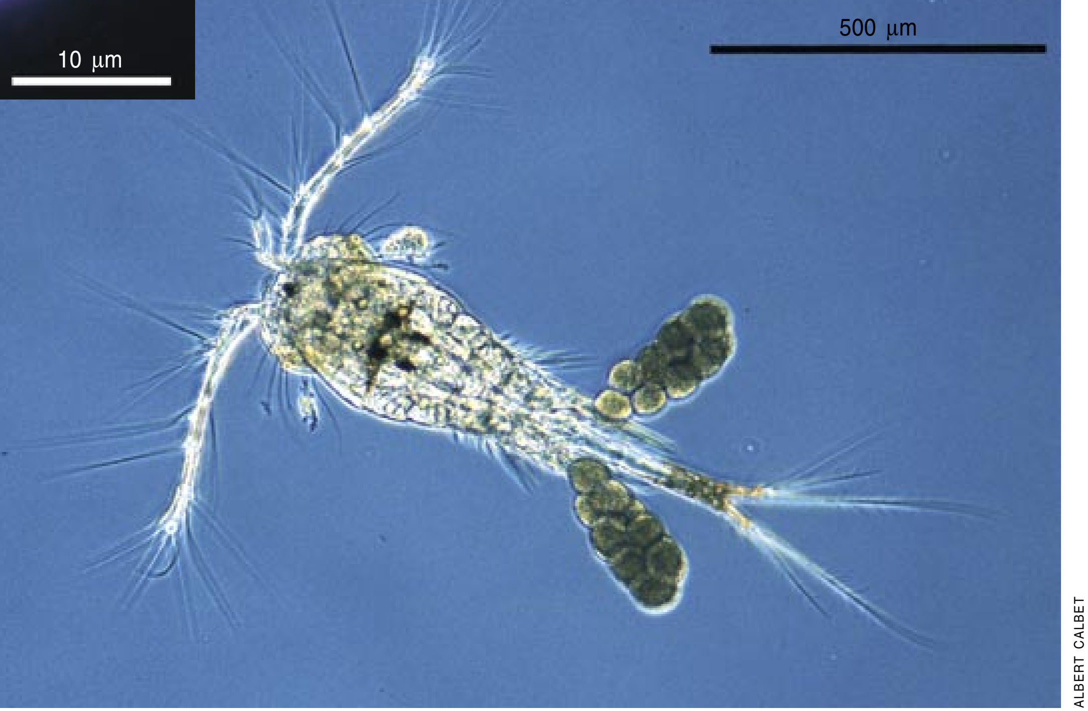
3. Muestra de fitoplancton recogida mediante una manga de nailon de 20 μm de malla. La fotografía al microscopio óptico permite observar la diversidad de formas y tamaños.



LAURA ARIN

4. *Alexandrium catenella*, dinoflagelado marino que puede provocar mareas tóxicas. Tinción con calcoflúor para poner de relieve las placas celulósicas que recubren el organismo. Los dinoflagelados son, en general, organismos sensibles a la turbulencia.

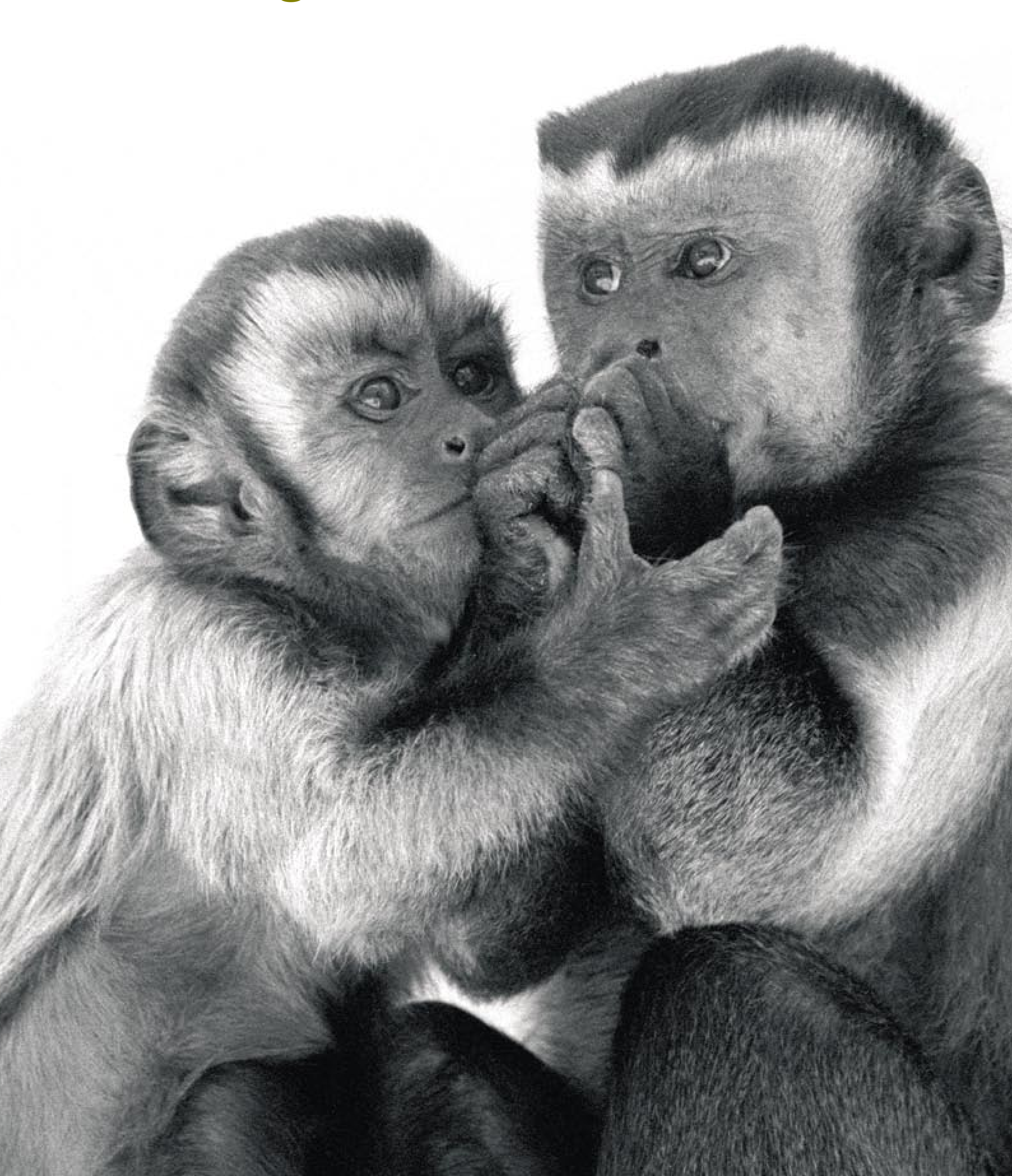
LAURA ARIN Y NAGORE SAMPEDRO



5. Copépodo de la especie *Oithona davisae*. Los copépodos son crustáceos pequeños que viven libres. Desempeñan una función clave en las redes tróficas marinas, pues sirven de puente alimenticio entre el plancton más pequeño y las larvas de peces y otros animales marinos.

ALBERT CALBET

Así negocian los animales



Frans B. M. de Waal

Los humanos y otros animales comparten un patrimonio de actitudes relacionadas con las transacciones: la cooperación, la devolución de favores y el resentimiento cuando se recibe menos de lo que se da

Lo mismo que mi despacho, que no tardaría en ser ocupado si yo me marchara a otra universidad, los predios silvestres no cesan de cambiar de manos. La oferta de hogares potenciales, muy rica, va desde agujeros taladrados por picos picapinos hasta conchas vacías en la playa. El mercado inmobiliario de los cangrejos ermitaños constituye un ejemplo típico de lo que los economistas denominan una "cadena de vacantes". Para proteger su blando abdomen, el cangrejo ermitaño acarrea su casa —por lo general una concha de gasterópodo abandonada— de un lado para otro. Con el tiempo, se ve obligado a cambiar de residencia: el cangrejo crece, pero la casa no. Por ello los cangrejos ermitaños están siempre al acecho, pendientes de encontrar un nuevo alojamiento. En el mismo instante en que uno se traslada a una concha más espaciosa, otros se ponen en fila, a la espera de ocupar la concha vacante.

Rige el proceso la ley de la oferta y la demanda. Sin embargo, dado el carácter "impersonal" de tales transacciones, muy pocos asociarían esa conducta con las transacciones económicas humanas. Las interacciones de los cangrejos ermitaños resultarían más interesantes si éstos llegaran a acuerdos del tipo "te cambio mi casa por este pez muerto". Pero los ermitaños no entienden de negociaciones, ni se andan en remilgos a la hora de echar al inquilino por la fuerza. Otros animales más sociales, en cambio, sí negocian. Su forma de intercambiar recursos y servicios arroja luz sobre cómo y por qué pudo evolucionar el comportamiento económico humano.

La nueva economía

La economía tradicional considera que las personas conceden la máxima importancia a los beneficios, movidas por puro egoísmo. Así lo indicaba Thomas Hobbes, filósofo inglés del siglo XVII: "Se da por sentado que los seres humanos buscan de modo natural lo que es bueno para ellos, y lo que es justo sólo por consideración de la paz, y accidentalmente". Según esta visión del comportamiento humano, todavía dominante, la sociabilidad constituye un añadido de última hora, un "contrato social" que nuestros antepasados aceptaron por el beneficio que reportaba, no porque sintieran una atracción mutua.

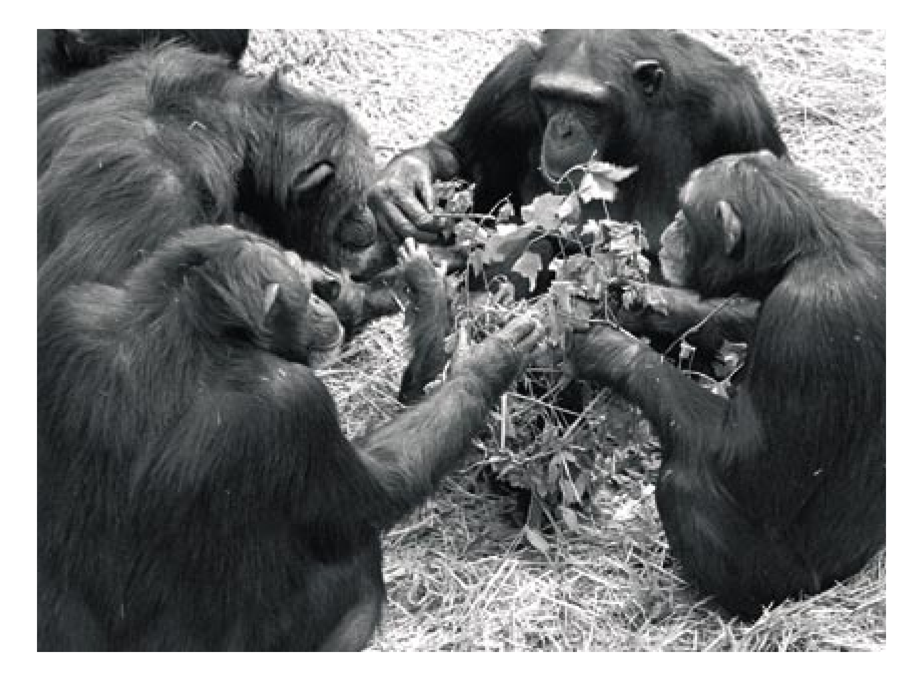
Para los biólogos, semejante exposición carece de fundamento. Descendemos de un extenso linaje de primates gregarios. Ello significa que la naturaleza nos ha dotado de un fuerte deseo de sentirnos integrados y de encontrar compañeros con los que vivir y trabajar. Esta explicación evolutiva de las relaciones sociales ha ido ganando aceptación con la aparición de una nueva disciplina: la economía del comportamiento, que estudia la toma de decisiones en el terreno de la economía, a partir del comportamiento humano real y no de las fuerzas abstractas del mercado. En 2002, la escuela recibió la sanción oficial con la concesión del Nobel de economía a dos de sus fundadores: Daniel Kahneman y Vernon I. Smith.

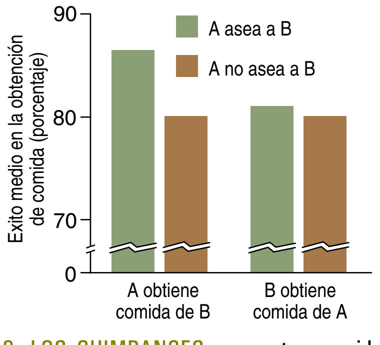
La incipiente economía del comportamiento animal demuestra que las tendencias y preocupaciones económicas básicas en humanos —reciprocidad, repartición de bienes y cooperación— no son exclusivas de nuestra especie. Probablemente aparecieron en otros animales a lo largo de la evolución por las mismas razones que se desarrollaron en nosotros: para que los individuos sacaran el máximo partido unos de otros sin socavar los intereses comunes que sostienen la vida en grupo.

A modo de botón de muestra, citaré un incidente que me ha sucedido en mi lugar de trabajo, el Centro Nacional Yerkes de Investigación de Primates en Atlanta. Habíamos enseñado a monos

FRANS B. M. DE WAAL

1. LOS MONOS CAPUCHINOS comparten la comida. Coinciden en ello con chimpancés y humanos. Rara entre otros primates, esta práctica podría haber evolucionado al par que la caza en grupo, estrategia utilizada por las tres especies. Sin recompensa compartida, no habría caza conjunta. En la fotografía, un mono capuchino juvenil ahueca la mano cerca del macho adulto para pedirle una ración de su comida.





2. LOS CHIMPANCES comparten comida (estas ramas con hojas, por ejemplo) a cambio de favores como el aseo. Tal reciprocidad se demostró experimentalmente observando que el acicalamiento mutuo de la mañana se correspondía con el acto de compartir comida por la tarde. El éxito del chimpancé A a la hora de obtener comida del chimpancé B aumentó después de que A hubiera aseado a B, pero el éxito de B a la hora de obtener comida de A no estaba afectado por el favor de A (*gráfica*). Es el aseador, pues, quien se beneficia, lo que ratifica la regla de trueque de comida por aseo.

capuchinos a alcanzar un tazón de comida situado sobre una bandeja; para ello sólo debían tirar de una barra fijada a la bandeja. Pesaba ésta demasiado para un solo sujeto; los monos tenían un motivo para trabajar juntos.

Les llegó el turno a Bias y Sammy, dos hembras. Situadas en jaulas contiguas, alcanzaron una bandeja en la que había dos tazones de comida. Pero Sammy tenía tanta prisa por obtener su recompensa que soltó la barra y agarró el tazón antes que Bias tuviera la oportunidad de asir el suyo. La bandeja salió disparada hacia atrás, quedando fuera del alcance de Bias. Mientras Sammy devoraba la comida, Bias se encrespó. Gritó con toda la fuerza de sus pulmones durante medio minuto, hasta que Sammy se acercó de nuevo a la barra de halar y ayudó a Bias a alcanzar la bandeja por segunda vez. Sammy no lo hizo para su propio provecho, pues para entonces su tazón ya estaba vacío.

Con esta enmienda de su comportamiento, Sammy respondía a la protesta de Bias ante la pérdida de una gratificación esperada. Semejante acción guarda con las transacciones económicas humanas un parecido mayor que la de los cangrejos ermitaños, pues muestra cooperación, comunicación, satisfacción de una expectativa y quizás un sentido de obligación. Sammy pareció basarse en un trueque, una reacción hasta cierto punto previsible, si tenemos en cuenta que la vida en grupo de los monos capuchinos gira alrededor de la misma mezcla de cooperación y competencia que caracteriza a nuestra propia sociedad.

La evolución de la reciprocidad

Lo mismo que las personas, los animales se prestan en ocasiones ayuda sin que medie un beneficio palmario para el que asiste. ¿Cómo se explica semejante altruismo? Cuando ello ocurre entre miembros de la misma familia, la respuesta es bastante sencilla: la sangre tira. Los biólogos reconocen ventajas genéticas en un comportamiento de este tipo: si tu pariente sobrevive, aumenta la probabilidad de que tus genes lleguen a la siguiente generación. La cooperación entre individuos no emparentados, en cambio, no sugiere ventajas genéticas inmediatas.

Pëtr Kropotkin, príncipe ruso, ofreció una primera explicación del fenómeno en su libro *Ayuda mutua*, publicado en 1902: si la ayuda es comunal, razonaba, todos los individuos ganan —la probabilidad de supervivencia aumenta para todos—. Pero tuvimos que esperar hasta 1971 para que Robert L. Trivers, entonces en la Universidad de Harvard, formulara esa idea en términos evolutivos modernos, con su teoría del altruismo recíproco.

En la opinión de Trivers, sacrificarse por otro sale a cuenta si, más tarde, éste devuelve el favor. La reciprocidad queda reducida a "te rasco la espalda si luego tu me rascas la

Resumen/Economía del comportamiento

- La teoría económica del comportamiento considera que nuestra forma de hacer negocios deriva de un legado que nuestra especie recogió en el curso de la evolución.
- La ley de la oferta y la demanda, así como el trueque, tan influyentes en el comercio de bienes y servicios en la economía de los humanos, afectan también a las actividades comerciales de los animales.
- La indignación ante tratos injustos y otras reacciones emocionales subyacen a las negociaciones entre humanos así como entre animales.
- Esta psicología compartida puede explicar comportamientos curiosos como el altruismo. Forman parte de nuestro bagaje de primates cooperadores.

FRANS B. M. DE WAAL (*fotografía*); ALISON KENDALL (*gráfica*)

mía". ¿Qué muestras de este *trueque* hallamos entre los animales? Monos y simios forman coaliciones; dos o más individuos, por ejemplo, atacan conjuntamente a un tercero. Asimismo, se ha encontrado una correlación positiva entre la frecuencia con que A apoya a B y la frecuencia con que B ayuda a A.

Pero, ¿significa ello que los animales llevan la cuenta de los favores dados y recibidos? Quizá simplemente dividan el mundo en "amigos", a los que prefieren, y "los demás", que les importan poco. Cuando el sentimiento es mutuo, la relación resulta recíprocamente fructífera o recíprocamente estéril. Tales simetrías explican la reciprocidad observada en peces, murciélagos vampiro (que regurgitan sangre a sus compañeros), delfines y diversas especies de monos.

Porque no anotan el debe y el haber de los favores, no hemos de suponer que carezcan del sentido de la reciprocidad. El planteamiento es otro. En concreto: ¿de qué modo encuentra el camino de retorno al altruista original un favor realizado? ¿En qué estriba exactamente el mecanismo de la reciprocidad? El registro mental de favores ofrece un posible mecanismo, pero está por saber si opera en los animales. Hasta la fecha, los chimpancés constituyen la única excepción. En estado salvaje, cazan en grupo monos colobos. Por lo general, un cazador se encarga de capturar la presa; luego la despedaza y la comparte. Pero sólo reciben ración aquellos que han tomado parte en la caza —aquel que no haya participado, aunque se trate del macho de mayor rango, puede esperar sentado, de nada le servirá pedir—. Este comportamiento sugiere reciprocidad en sí mismo; durante el reparto de los despojos, gozan de prioridad los participantes en la cacería.

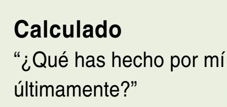
Para conocer qué mecanismos operan, aprovechamos la tendencia de esos primates a compartir, una conducta observada también en cautividad. Ofrecimos a uno de los chimpancés de nuestra colonia una sandía o algunas ramas con hojas. Al principio, el propietario se hallaba en el centro de un grupo de participación; pronto aparecerían grupos secundarios alrededor de individuos que habrían conseguido hacerse con una buena ración, hasta que la comida se hubiera repartido entre todos.

Entre los chimpancés rige el "respeto de posesión"; no suele verse que ninguno exija por la fuerza la comida de otro. Los pedigüeños extienden la mano, con la palma hacia arriba, a la manera de nuestros mendigos. Lloriquean y gimotean, pero no recurren a la violencia. Si se entabla por casualidad una pelea, suele iniciarla el poseedor con el fin de que otro abandone el grupo. Ataca a los transgresores golpeándoles la cabeza con una rama gruesa o les grita con voz chillona hasta que lo dejan solo. Sea cual sea su rango, los poseedores controlan el reparto compartido de la pitanza.

Analizamos unas 7000 situaciones de este tipo. Estudiamos la tolerancia que el poseedor mostraba hacia pedigüeños concretos en función de los servicios recibidos previamente. La mañana de los días en que estaba

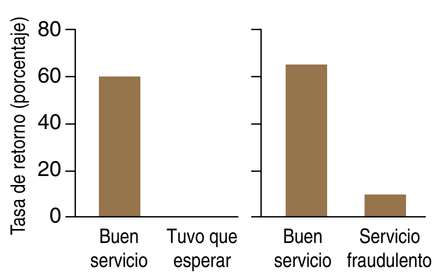
Cómo opera la reciprocidad

Los humanos y otros animales intercambian bienes a través de distintos mecanismos de reciprocidad que comparten un rasgo común: el beneficio repercute luego en el donante de partida.

MECANISMO DE RECIPROCIDAD	RASGOS FUNDAMENTALES
Basado en la simetría "Porque somos amigos" 	El afecto mutuo entre dos individuos promueve un comportamiento similar en ambas direcciones, sin necesidad de llevar un registro mental diario del balance de favores, mientras la relación resulte, en general, satisfactoria. Posiblemente sea éste el mecanismo de reciprocidad más común en la naturaleza; define relaciones estrechas lo mismo entre humanos que entre chimpancés. **Ejemplo:** Los chimpancés "amigos" se asocian, se acicalan juntos y se ayudan los unos a los otros en las peleas.
De actitud "Si eres amable conmigo, yo lo seré contigo" 	Cada individuo imita las actitudes del otro, intercambiando favores sobre el terreno. Esta reciprocidad se manifiesta de forma instantánea entre monos. Suele guiar las relaciones entre personas que se conocen poco. **Ejemplo:** Un mono capuchino comparte comida con el que le ayuda a obtenerla (*véase el ejemplo de la bandeja*).
Calculado  "¿Qué has hecho por mí últimamente?"	Los individuos llevan la cuenta de los favores que intercambian con cada compañero; ello les ayuda a decidir a quién devolverlos. Este mecanismo es típico de los chimpancés y común en las relaciones distantes y profesionales entre personas. **Ejemplo:** Un chimpancé que haya aseado a un compañero por la mañana, esperará que éste le recompense por la tarde con comida.

ROBERTO OSTI





3. LOS PECES LIMPIADORES comen los parásitos de la boca de sus clientes, de mayor tamaño. Los clientes divagantes raramente retornan a la “barbería” de un pez limpiador, si se les ha hecho esperar *(histograma de la izquierda)* o si el limpiador ha mordido tejido sano del cliente *(histograma de la derecha)*. Por ello los peces limpiadores tienden a tratar mejor a los clientes divagantes que a los residentes, que no tienen otras barberías a su alcance.

planeada una prueba de compartir comida, registrábamos los detalles del acicalamiento. Si el macho situado en la cima de la jerarquía, Socko, había acicalado a May, por ejemplo, aumentaba la probabilidad de obtener de ella unas cuantas ramas por la tarde. Esta relación entre comportamiento pasado y presente resultó ser general. Pero el patrón variaba de un día al siguiente; por tanto, no se trataba de conexiones simétricas. El nuestro fue el primer estudio con animales que demostraba una correlación entre los favores dados y los recibidos. El trueque de “comida por acicalamiento” era, asimismo, específico del compañero: es decir, la tolerancia de May beneficiaba a Socko, que la había acicalado, pero a ningún otro.

Este mecanismo de reciprocidad requiere la memoria de acontecimientos anteriores y la coloración de la misma con tintes amistosos. En nuestra propia especie, este proceso de coloración se denomina gratitud; no hay razón para llamarlo de otra manera en los chimpancés. Ignoramos si los primates sienten también obligaciones, aunque encierra notable interés que la tendencia a devolver favores no sea la misma en todos los tipos de relaciones. Entre individuos que mantienen una relación estrecha —los que se asocian y se acicalan de forma asidua—, el peso de una sola sesión de aseo resulta insignificante. Se dan entre ellos todo tipo de intercambios cotidianos, probablemente sin que lleven la cuenta de los mismos. Parece operar, en cambio, el criterio de amistad comentado antes. Sólo en las relaciones más distantes, el acicalamiento destaca como merecedor de recompensa. Puesto que Socko y May no eran “amigos íntimos”, el acicalamiento de Socko fue debidamente advertido.

Un patrón similar se observa en el comportamiento humano: nos sentimos más inclinados a llevar la cuenta de favores con los extraños y compañeros que con nuestros amigos y familiares. Es más, llevar el control de los favores entre las personas más próximas —entre marido y mujer, por ejemplo— constituye una clara señal de desconfianza.

Mercados biológicos

Puesto que la reciprocidad es cosa de dos, la elección de pareja, o “socio”, constituye una cuestión central en la economía del comportamiento. El intercambio de viviendas de segunda mano entre cangrejos ermitaños es de extrema sencillez si se compara con las transacciones entre primates, en las que múltiples individuos participan en un trueque de múltiples bienes y servicios: acicalamiento, sexo, apoyo en las luchas, comida y cuidado de las crías, entre otras.

En ese “mercado de servicios”, conviene mantener buenas relaciones con los que ocupan un rango superior en la jerarquía para fomentar las asociaciones de acicalamiento y —si



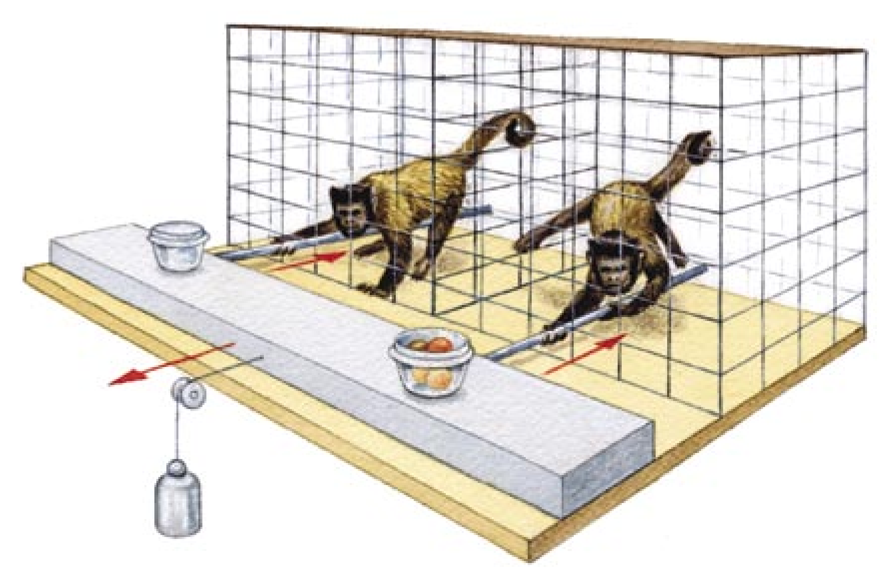
4. LAS HEMBRAS DE PAPION pagan un precio en acicalamiento para poder mirar a una cría a hurtadillas. A mayor escasez de crías, mayor es el tiempo de aseo que se paga para observarlas. El precio de la mercancía (la cría de papión, en este caso) aumenta a medida que cae la oferta (hay menos crías).



se es ambicioso— cerrar tratos con otros del mismo parecer. Los machos de chimpancé, por ejemplo, forman coaliciones para desafiar al macho dominante. Se trata de una empresa de alto riesgo, pues, tras el derrocamiento, el nuevo mandatario debe complacer a los que lo apoyaron. De hecho, lo más probable para un macho alfa que pretenda monopolizar los privilegios del poder, como el acceso a las hembras, es que su posición le sea arrebatada en poco tiempo. No necesitan haber leído a Maquiavelo.

Al "comprar" cada individuo los mejores socios y "vender" sus propios servicios, el entramado de reciprocidad se convierte en un sistema de oferta y demanda, que se corresponde con el concepto que Ronald Noë y Peter Hammerstein, a la sazón en el Instituto Max Planck de Fisiología del Comportamiento en Seewiesen, desarrollaron en su teoría del mercado biológico. Esta teoría, aplicable siempre que exista libre elección de los socios mercantiles, postula que el precio de las mercancías y de los socios varía en función de su disponibilidad. Dos estudios sobre las fuerzas del mercado ilustran este punto: uno se refiere al mercado de las crías de papiones; el otro, al servicio que ofrecen los peces limpiadores.

Como todas las hembras de primates, las de papiones se sienten irresistiblemente atraídas hacia las crías, sean o no progenie suya. Emiten simpáticos gruñidos e intentan tocarlas. Pero las madres son muy protectoras; se muestran reacias a dejar que nadie manosee a sus preciosos recién nacidos. Para acercarse, la hembra interesada acicala a la madre al propio tiempo que atisba a la cría por encima de la espalda o por debajo del brazo de la progenitora. Tras una relajante sesión de acicalamiento, la madre puede acceder al deseo de la aseadora de observar la cría más de cerca. Así se compra el "tiempo de cría". La teoría de mercado predice que el valor de ese tiempo aumenta cuando las crías escasean. En un estudio de papiones chacma salvajes en Sudáfrica, Louise Barrett, de la Universidad de Liverpool, y Peter Henzi, de la Universidad de Lancashire Central, encontraron que, en efecto, las madres de crías poco numerosas podían obtener mayores



5. UN EXPERIMENTO en el que debe tirarse de una bandeja para obtener comida demuestra que los monos capuchinos tienen más probabilidades de compartir alimento con socios cooperadores que con los que no colaboran. La jaula del ensayo alberga dos capuchinos, separados por una reja. Enfrente se les coloca una bandeja que pesa demasiado (dotada de un contrapeso) para que un mono consiga moverla él solo. Para arrastrarla, pues, deben colaborar. Pero al tirar de la barra sólo uno de ellos (el "ganador") obtiene recompensa: un tazón lleno de comida. El tazón (transparente) del otro mono (el "bracero"), en cambio, está vacío; por tanto, éste trabaja sólo para el "ganador", quien suele compartir la comida con el bracero a través de la reja. De lo contrario, el bracero perderá interés en ayudarle en próximas ocasiones.

ROBERTO OSTI

beneficios (acicalamientos más prolongados) de la "venta" del tiempo de su cría que las madres de un grupo con un gran número de crías.

Las doncellas limpiadoras *(Labroides dimidiatus)* son pececillos marinos que se alimentan de los parásitos externos de peces mayores. Cada limpiador posee una estación de limpieza o "barbería" en un arrecife, a la que acuden los clientes. Estos extienden sus aletas pectorales y adoptan posturas que facilitan la tarea del limpiador. El intercambio ejemplifica un perfecto mutualismo.

El "barbero" mordisquea los parásitos y los arranca de la superficie del cuerpo del cliente, de sus branquias e incluso del interior de su boca. En ocasiones, no da abasto; está tan ocupado que los clientes deben esperar haciendo cola. Entre la clientela se distinguen dos tipos de peces: los residentes y los divagantes. Los residentes pertenecen a especies con territorios reducidos; no tienen otra elección que acudir al limpiador local. Los divagantes, en cambio, o bien medran en territorios extensos o realizan largos viajes, lo que significa que pueden elegir entre varias estaciones de limpieza. Del servicio valoran la rapidez, la calidad y también la honradez. El fraude se produce si el limpiador arranca un bocado de mucus sano. Cuando ello sucede, el cliente, estremecido, abandona el establecimiento de inmediato.

La investigación sobre las doncellas limpiadoras realizada por Redouan Bshary, del Instituto Max Planck de Seewisen, no se circunscribe a observaciones *in situ*, en el arrecife; abarca también ingeniosos experimentos en el laboratorio. Sus artículos parecen un manual de buenas prácticas comerciales. Los clientes divagantes se muestran más propensos a cambiar de "barbería" si un limpiador los ha hecho esperar demasiado tiempo o los ha engañado. Los limpiadores parecen saberlo; tratan mejor a los divagantes que a los residentes: si un divagante y un residente llegan al mismo tiempo, el limpiador casi siempre atiende primero al divagante; sin otro lugar adonde acudir, los residentes pueden esperar. La única categoría de peces a los que los limpiadores no engañan nunca son los depredadores, pues cuentan con una contraestrategia radical: si no quedan satisfechos con el servicio, se tragan al "barbero". Con los depredadores, los peces limpiadores, prudentes, adoptan, en palabras

de Bshary, una "estrategia incondicionalmente cooperativa".

La teoría del mercado biológico ofrece una elegante solución al problema de los "gorrones". Los vividores —individuos que viven a expensas de los demás— alteran fácilmente los sistemas de reciprocidad. Los teóricos suelen asumir que la mejor solución es castigar al estafador, pero eso no ha podido demostrarse en el mundo animal. Existe otra forma, más simple, de resolverlo: en una situación de libre elección de socios, los animales simplemente abandonan las relaciones insatisfactorias y las sustituyen por las que ofrecen beneficios. Los mecanismos de mercado son todo lo que se necesita para evitar que los aprovechados se salgan con la suya. Tampoco en nuestras sociedades nos causan simpatía los que toman más de lo que dan, ni nos merecen confianza. Tendemos a alejarnos de ellos.

Un trato es un trato

Para beneficiarse de la cooperación, un individuo debe comparar su esfuerzo con el esfuerzo de otros y sus recompensas con el esfuerzo invertido. Para averiguar si los animales llevan a cabo este seguimiento, volvimos a nuestros monos capuchinos. Simulamos un mercado de trabajo (en miniatura) inspirado en el comportamiento que muestran los capuchinos cuando atacan ardillas gigantes. Aunque la caza de las ardillas es una empresa colectiva, sólo un individuo captura la presa y en sus manos recae entera la recompensa. Ahora bien, si el captor se quedara con el premio, no cuesta imaginar que la próxima vez habría de ingeniárselas solo. Los capuchinos comparten la carne por la misma razón que los chimpancés (y las personas): no hay cacería conjunta sin retribuciones conjuntas.

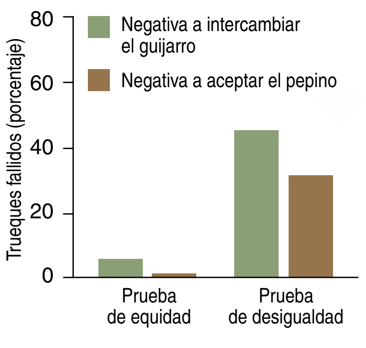
Reprodujimos esta situación en el laboratorio asegurándonos de que sólo un mono (al que denominamos ganador) de la pareja que tiraba de una bandeja recibía un tazón con trozos de manzana. El socio (el bracero), en cambio, no tenía comida en su tazón, lo que resultaba evidente desde el principio porque los tazones eran transparentes. El bracero tiraba, pues, para beneficio del ganador. Los monos se hallaban situados uno junto al otro, separados por una reja. De pruebas anteriores sabíamos que los poseedores de comida podían compartirla, permitiendo que el vecino la alcanzara a través de la reja. En raras ocasiones, empujaban los fragmentos de comida hacia el otro.

Comparamos los tirones de bandeja colectivos con los tirones individuales. En una situación, ambos animales contaban con una barra para arrastrar una bandeja pesada; en la otra, el bracero carecía de barra y el ganador manejaba una bandeja más ligera él solo. Se produjo un mayor número de compartición del alimento tras arrastres conjuntos que tras tirones solitarios. El ganador compensaba a su socio por la ayuda prestada. Confirmamos también que compartir afecta a la cooperación futura. El pago del bracero constituye una estrategia inteligente si tenemos en cuenta que la tasa de éxito de una pareja cae cuando el ganador no reparte los beneficios de la colaboración.

Sarah F. Brosnan, colaboradora mía en Yerkes, llevó un paso más adelante la exploración de las reacciones ante las distintas formas de repartir las recompensas. Ofrecía a un mono capuchino un guijarro; después le mostraba una tajada de pepino como aliciente para que devolviera el guijarro. Los monos comprendieron rápidamente el principio de intercambio. Situados uno junto al otro, dos monos intercambiaban de buen grado guijarros por pepinos con la investigadora. Pero, si a uno de ellos se le daban uvas mientras que el otro seguía recibiendo pepinos, las cosas tomaban un giro inesperado. (Las uvas constituyen uno de sus manjares favoritos.) Los monos que hasta entonces se habían mostrado dispuestos a seguir el juego para obtener pepino, de repente se pusieron en huelga. No sólo actuaban de mala gana al ver que el otro conseguía mejor trato, sino que se agitaban y lanzaban los guijarros fuera de la cámara de prueba y, en ocasiones, también las tajadas de pepino. Un alimento que en condiciones normales jamás se hubiera rechazado, se había tornado despreciable.

Rechazar una paga desigual —algo que hacen también las personas— va contra las premisas de la economía tradicional. Si maximizar los beneficios fuera lo único que importara, uno debería tomar todo lo que estuviera a su alcance y no permitir nunca que el resentimiento o la envidia interfirieran. Los economistas del comportamiento suponen que la evolución ha llevado a emociones que retienen el espíritu de cooperación y que influyen sobremanera en el comportamiento. A corto plazo, preocuparse de lo que los

FRANS B. M. DE WAAL; ALISON KENDALL *(gráfica)*





6. LOS MONOS CAPUCHINOS tienen sus preferencias cuando se trata de comida. Anteponen la fruta a las hortalizas, como el apio que este capuchino come con aire pensativo, por ejemplo. Se les enseñó a una pareja de monos capuchinos a intercambiar un guijarro por una tajada de pepino. Así lo hicieron de buena gana mientras a ambos se les recompensaba con el mismo premio: pepino *(prueba de equidad)*. Pero cuando a uno de los dos se le empezó a premiar con un racimo de uvas, el que continuaba recibiendo pepino se negó a seguir el "juego sucio" *(prueba de desigualdad)*. Rehusó aceptar el pepino; en ocasiones, llegó incluso a echarlo fuera de la jaula o bien se negó a devolver el guijarro.

Así negocian los humanos

Las emociones que Frans de Waal describe en los intercambios que realizan los animales sociales mantienen paralelismos con nuestras propias transacciones económicas. Dichas similitudes sugieren que las interacciones económicas humanas están controladas, al menos en parte, por tendencias y emociones ancestrales. Los experimentos con animales constituyen la base de una floreciente disciplina: la economía del comportamiento. Esta pone en tela de juicio y modifica el modelo tradicional de la investigación económica, que sostiene que los humanos basan las decisiones económicas en procesos de pensamiento racionales. Por ejemplo, la gente rechaza ofertas que considera injustas, mientras que, según la economía clásica, uno debería tomar todo lo que esté a su alcance. En 2002, el premio Nobel de economía se concedió a dos pioneros de este nuevo campo: Daniel Kahneman, psicólogo de la Universidad de Princeton, y Vernon L. Smith, economista de la Universidad George Mason.

Kahneman, con su colega Amos Tversky (inelegible para el premio porque murió en 1996) analizaron en los humanos la toma de decisiones en un contexto de incertidumbre y riesgo. Los economistas tradicionales describían las decisiones humanas en términos de expectativas de utilidad: la suma de las ganancias que uno piensa que obtendrá de algún acontecimiento futuro multiplicada por la probabilidad de que éste ocurra. Pero Kahneman y Tversky demostraron que el temor a las posibles pérdidas suele superar la ilusión por las ganancias potenciales, con lo que la mayoría opta por mostrar un comportamiento gregario. El estallido de la burbuja del mercado de valores en 2000 ofrece un poderoso ejemplo de este comportamiento: el deseo de permanecer con el rebaño quizás hizo que la gente desembolsara mucho más por las acciones de lo que habría pagado cualquier inversor racional.



UN ENTUSIASMO IRRACIONAL se apodera del parquet de una bolsa de valores.

El trabajo de Smith demostró, asimismo, que los experimentos de laboratorio también funcionan en economía, una ciencia tradicionalmente considerada no experimental, basada sólo en la observación. Entre sus hallazgos en el laboratorio: las decisiones emocionales no siempre son desaconsejables.

—La redacción

demás obtienen puede parecer irracional, pero a la larga permite evitar que se aprovechen de uno. Desalentar la explotación resulta fundamental para persistir en la cooperación.

Con todo, estar siempre pendiente del flujo de beneficios y favores resulta bastante tedioso. Por eso los humanos se protegen contra los que viven de gorra y los explotadores estableciendo lazos de amistad con los "socios" que han resistido el paso del tiempo: el cónyuge o un buen amigo, por ejemplo. Una vez hemos determinado de quién podemos fiarnos, relajamos las reglas. Sólo con socios más distantes mantenemos un registro mental del balance de favores y reaccionamos enérgicamente ante los desequilibrios, que consideramos "injustos".

El efecto de la distancia social se pone de manifiesto también entre los chimpancés. Como hemos visto, el *trueque* estricto es raro entre amigos que se hacen favores de manera rutinaria. Dichas relaciones parecen inmunes a la inequidad. Brosnan realizó su prueba de trueque con uvas y pepinos en chimpancés. La reacción más fuerte se produjo entre los individuos que se conocían desde hacía poco tiempo; en cambio, los miembros de una colonia que habían vivido juntos durante más de 30 años apenas reaccionaron. Posiblemente, cuanto mayor es la familiaridad, más extensa es la ventana de tiempo en el transcurso del cual los chimpancés evalúan su relación. Sólo las relaciones distantes se muestran sensibles a las fluctuaciones del día a día.

Todos los agentes económicos, humanos o animales, deben ocuparse del problema de los "gorrones" y del reparto de los beneficios de una empresa colectiva. Para ello comparten todo lo posible con los que más los ayudan y exhiben fuertes reacciones emocionales ante la violación de las expectativas. La teoría económica evolutiva reconoce esta psicología compartida y considera la sociabilidad no un fenómeno accidental, Hobbes *dixit*, sino parte de nuestro bagaje en cuanto primates cooperadores.

JIM SUGAR *Corbis*

El autor

Frans B. M. de Waal ocupa la cátedra C. H. Candler de etología de los primates en la Universidad de Emory donde dirige el Centro de Vínculos Vivos del Instituto Nacional Yerkes de Investigación de Primates. Se ha especializado en el comportamiento y cognición sociales de monos, chimpancés y bonobos; sobre todo en la cooperación, la resolución de conflictos y la cultura.

Bibliografía complementaria

PAYMENT FOR LABOUR IN MONKEYS. Frans B. M. de Waal y Michelle L. Berger en *Nature*, vol. 404, pág. 563; 6 abril, 2000.

CHOOSY REEF FISH SELECT CLEANER FISH THAT PROVIDE HIGH-QUALITY SERVICE. R. Bshary y D. Schäffer en *Animal Behaviour*, vol. 63, n.º 3, págs. 557-564; marzo, 2002.

INFANTS AS A COMMODITY IN A BABOON MARKET. S. P. Henzi y L. Barrett en *Animal Behaviour*, vol. 63, n.º 5, págs. 915-921; 2002.

MONKEYS REJECT UNEQUAL PAY. Sarah F. Brosnan and Frans B. M. de Waal en *Nature*, vol. 425, págs. 297-299; 18 septiembre, 2003.

El origen de la materia

Aún no sabemos por qué la materia dominó a la antimateria en la formación del universo

James M. Cline

¿De dónde proviene la materia? Quizá no deje perplejos a muchos que existamos y nos rodee la materia, que no sea el mundo un mero baño de microondas. Mas, para un cosmólogo, la existencia de materia es desconcertante, un problema que no ha encontrado solución desde que la física teórica hubo de planteárselo hace casi cuarenta años. Las mejores teorías acerca del origen del universo no ofrecen todavía una explicación.

La existencia de materia es un capítulo inacabado de la teoría de la gran explosión del origen del universo, que por lo demás acierta a explicar casi todo lo que observamos. Se cree que la materia y la antimateria se equilibraban exactamente entre sí en los primeros momentos tras la gran explosión. Como la materia y la antimateria no pueden coexistir —se aniquilan la una a la otra—, esta situación, si hubiera continuado, habría dejado un universo muy monótono. ¿Por qué prevaleció la materia y llegó a dominar el universo?

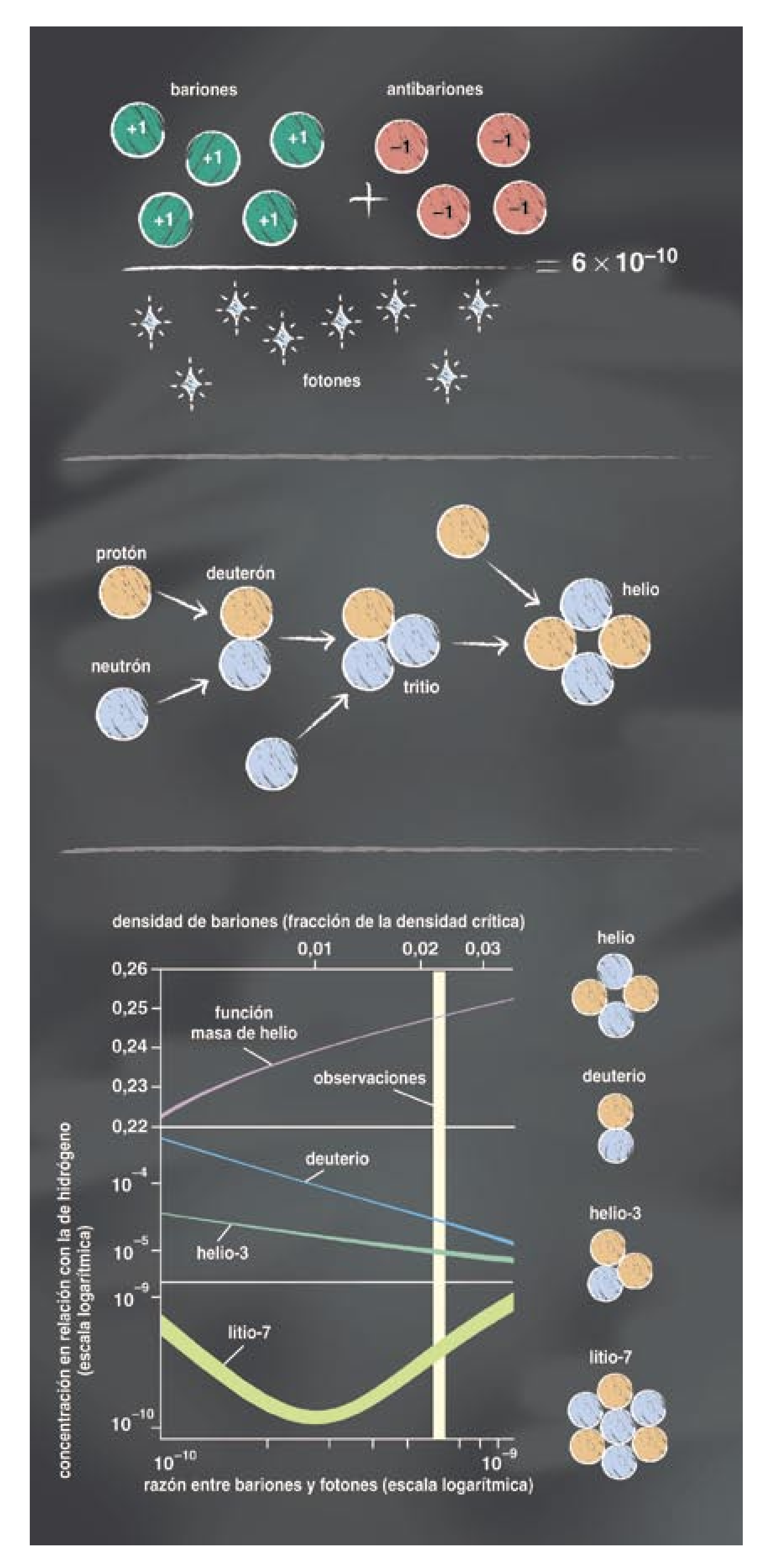
El físico y disidente político ruso Andrei Sajarov se adelantó a su tiempo al proponer ideas relativas a la creación de la materia en 1967. Por entonces, cuando yo estaba aún aprendiendo la tabla de multiplicar, el origen de la materia se daba más o menos por descontado. Sajarov vio la importancia de la cuestión. Considerando que la creación de la materia no podía darse por descontada, enunció las condiciones que una explicación plausible debería satisfacer.

La materia de que estamos hechos consta sobre todo de protones y neutrones, los componentes de los núcleos atómicos. Se llama a estas partículas *bariones*, de la palabra griega *barys*, que quiere decir pesado: lo son 2000 veces más que los electrones. Los experimentos han demostrado que se componen de partículas menores, los quarks. No faltan bariones en el universo. Su densidad media es de unos 0,2 por metro cúbico. Pero en ese mismo volumen hay *diez mil millones* de fotones, o partículas de luz. Se trata de los fotones del fondo cósmico de microondas, el remanente de la gran explosión.

La razón entre la densidad de bariones y la de fotones permanece constante a lo largo del tiempo, pese a que el universo se va expandiendo y diluyendo. La *asimetría bariónica* del universo, el número de bariones dividido por el número de fotones, es 6×10^{-10}. Un número extraño, que nunca ha dejado de preocupar a los físicos teóricos. Si la razón hubiera sido más próxima a 1, hubiéramos podido imaginar que nada muy especial sucedió en el universo primitivo, ya que todas las partículas se comportan más o menos igual a temperaturas elevadísimas; entonces se producen bariones y fotones en números aproximadamente iguales. También hubiera resultado natural que no hubiese habido ni un barión: así habría ocurrido si la materia y la antimateria se hubiesen aniquilado mutuamente. Al calcular la asimetría bariónica, se sustraerían los antibariones y saldría cero.

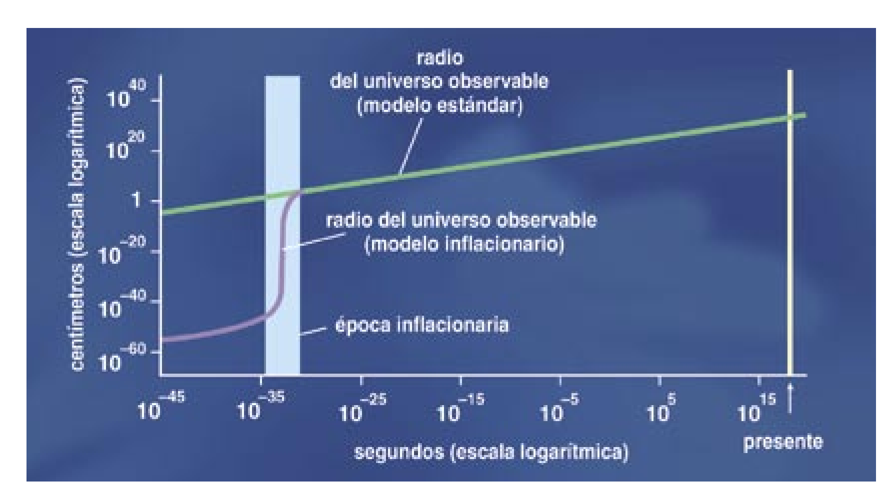
Claro que entonces no habría físicos que perdiesen el sueño por estas cosas: no habría en el universo más que fotones de microondas. Pero un universo sin bariones debe considerarse una posibilidad natural, ya que por todas las partes de la física encontramos principios de simetría. Un tal principio existe para la carga eléctrica: sabemos que en el universo, por cada partícula cargada positivamente, debe haber una negativa.

En contraste, el universo bariónico parece exigir que exista una asimetría. La magnitud de la misma debería seguirse de las ecuaciones de una teoría de la bariogénesis, del origen de la materia. Desgraciadamente, no parece que en lo que nos quede de vida vaya a resultar posible someter a prueba con experimentos la multitud de teorías formuladas al respecto. No podemos esperar que las circunstancias imperantes en los primerísimos momentos del universo, los más calientes, vayan a reproducirse en ningún laboratorio. ¿Es esta área de la teoría un simple juego, en el que podemos construir hábiles hipótesis que sólo juzgaremos por su coherencia y elegancia? Quizá; pero los experimentadores ya han acumulado datos que descartan algunas propuestas, incluida una en la que estuve trabajando cinco años. Como las nuevas teorías siempre nacen para ser sometidas a comprobación, podemos mantener la esperanza de que las ideas que ahora se desarrollan —incluyendo la posibilidad de que la asimetría bariónica surja de una asimetría entre partículas más ligeras, relacionadas con los electrones— den a los experimentadores ideas para otras pruebas. Nuevos hallazgos, o nuevos enfoques, podrían dar nueva



1. LA ASIMETRIA BARIONICA es la razón entre los números de bariones (neutrones y protones) y de fotones (*arriba*). A los bariones de la antimateria se les da un valor negativo en esa relación. Se sabe que la asimetría bariónica del universo es de 6×10^{-10}, un número que no puede explicarse en el marco de las teorías hoy aceptadas pese a que la observación lo ha dejado bien establecido. El número de bariones se determina por la concentración relativa, en el universo, de los elementos ligeros, que se formaron cuando los bariones se unieron y crearon núcleos atómicos. Esta etapa primitiva de la evolución del universo recibe la denominación de nucleosíntesis de la gran explosión (*medio*). La gráfica inferior muestra que las predicciones teóricas sobre las proporciones relativas de los elementos ligeros dependen de la asimetría bariónica. La banda vertical indica el valor de la asimetría en el que las predicciones concuerdan con las observaciones.

EMMA SKURNICK/*American Scientist*



2. LA INFLACION ES UNA VARIANTE de la teoría de la gran explosión del universo, corroborada por las últimas mediciones del fondo cósmico de microondas. Consistió en un período de expansión exponencialmente rápida en los primeros momentos del universo; en su transcurso, la densidad de bariones cayó. Para que la asimetría barión-antibarión se hubiese creado antes de la fase final de la inflación, o "recalentamiento", su valor inicial tendría que haber sido enorme.

vida a ideas ahora relegadas. En cualquier caso, a lo largo de los próximos años las mediciones astronómicas y los experimentos con aceleradores de partículas nos aportarán sin duda una abundancia de datos e ideas acerca del universo primitivo. Cabe esperar que aprenderemos mucho cuando comparemos nuestras teorías con esos resultados.

La génesis de la bariogénesis

Andrei Sajarov se pasó veinte años diseñando armas nucleares para la Unión Soviética. Pero poco a poco se convirtió en uno de los más duros críticos del régimen comunista. En 1980 lo confinaron. Mientras sus opiniones políticas le reportaban titulares de prensa en el extranjero y complicaciones en Moscú, fue abriendo nuevos caminos en cosmología. Entre sus extraordinarios logros se cuenta el inesperado hallazgo de los ingredientes necesarios para la creación de una asimetría bariónica. He dicho "inesperado" porque, en aquel tiempo, nadie más pensaba en tal dirección. Los proponentes de la teoría de la gran explosión suponían que el contenido bariónico del universo era una condición inicial que se debía imponer, un punto de partida y no un hecho que una teoría completa de la evolución cosmológica tuviese que explicar. La mayoría de los colegas de Sajarov en el Instituto Lebedev de Física de Moscú no apreciaron en su momento la importancia de su descubrimiento. Sin embargo, hoy día los cosmólogos saben que acertó.

Sajarov había estado reflexionando acerca de la violación de CP, una asimetría entre ciertas partículas y sus antipartículas. Esta asimetría, que será importante en lo que sigue, fue uno de los múltiples descubrimientos de física experimental de partículas, por entonces en auge. Se hacían chocar haces de partículas en aceleradores cada vez más grandes y potentes a fin de descubrir nuevos fenómenos que no se habían visto en las moderadas condiciones que prevalecen en la superficie de la Tierra. Los teóricos disponían también de alguna información útil acerca de las propiedades del propio universo.

Las concentraciones relativas de los elementos ligeros en el universo eran uno de esos datos. Los elementos ligeros —el helio, el litio y el deuterio, isótopo del hidrógeno— se formaron a los pocos minutos de la gran explosión, en la etapa que recibe el nombre de "nucleosíntesis de la gran explosión". A medida que el universo se enfrió tras la gran explosión, los protones y los neutrones se fueron uniendo y formaron los núcleos atómicos. Las colisiones con los fotones de alta energía tendían a romper estos núcleos, pero gradualmente los fotones perdieron energía debida al enfriamiento causado por la expansión del universo. No sorprende demasiado que la eficiencia con la que se producían los núcleos dependiese de la razón entre bariones y fotones. Uno de los principales triunfos de la teoría de la gran explosión consiste en que predice correctamente, una vez se ha elegido un valor para la asimetría bariónica del universo, las proporciones de todos los elementos ligeros que se generaron durante la nucleosíntesis de la gran explosión.

En 2003, nuevas mediciones precisas de las pequeñas fluctuaciones de la temperatura de la radiación cósmica de microondas confirmaron el valor conocido de la asimetría bariónica. Estas fluctuaciones revelan finos detalles de las ondas sonoras (o de presión) que existieron en la época en que los electrones se estaban combinando con los núcleos atómicos para formar átomos. Al mismo tiempo, el universo se volvió transparente para los fotones, liberados de la prisión del plasma caliente que dominó los primeros 100.000 años después de la nucleosíntesis. Este es otro éxito impresionante de la cosmología de la gran explosión: la asimetría bariónica, confirmada por dos mediciones del todo diferentes.

Pero los físicos encuentran flancos débiles en cualquier argumento con apariencia de solidez. Podemos imaginar, por ejemplo, que hay tantos antibariones como bariones a distancias muy grandes; nosotros viviríamos en una región dominada por bariones. De esta manera, el verdadero número bariónico sería cero. Las regiones de antimateria deberían estar más allá del límite del universo visible; si existieran galaxias de antimateria en la parte del universo que podemos ver, deberíamos detectar regiones fronterizas donde las aniquilaciones entre materia y antimateria produjesen fotones muy energéticos, o sea, rayos gamma. No observamos algo así, pero tampoco se conoce una manera verosímil de segregar en el universo primitivo la materia de la antimateria en zonas muy extensas. Resulta mucho más fácil inventar teorías de bariogénesis que de la segregación de la materia y la antimateria.

¿Hasta qué punto se necesita una explicación de la asimetría bariónica? La cifra 6×10^{-10} es tan buena como cualquier otra. ¿Y si no es más que un valor inicial aleatorio que surgió

EMMA SKURNICK/*American Scientist*

del caos de la gran explosión, tal y como pensaban los primeros teóricos de la gran explosión? Sin embargo, abundan los indicios de que es necesario un mecanismo especial para la bariogénesis.

Así, hay que tener muy en cuenta las poderosas razones que llevan a creer en la existencia de un período de expansión exponencialmente rápida justo tras la gran explosión, o período de "inflación", durante el cual la temperatura del universo fue a todos los efectos nula. La teoría de la inflación, creada a principios de los años ochenta, por Alan Guth, Andrei Linde, Paul Steinhardt y Andreas Albrecht, es una modificación de la teoría de la gran explosión. La inflación acaba con un "recalentamiento", proceso en el que la energía potencial almacenada en el vacío se convierte en partículas calientes con una cierta temperatura inicial finita. (En el modelo estándar de la gran explosión, la temperatura crece arbitrariamente a medida que se retrocede hacia el inicio del tiempo.) Las nuevas mediciones del fondo cósmico de microondas aportan un valioso respaldo a esa hipótesis. La inflación explica por qué las fluctuaciones de temperatura del fondo de microondas son pequeñas (del orden del 0,001 por ciento), pero no exactamente cero: fluctuaciones cuánticas de este tamaño se producen de manera natural durante la inflación.

En el curso de la inflación, los bariones se diluyen en el volumen del universo inflado. Si se supone que no se creó una asimetría barión-antibarión durante o después del recalentamiento, se deberá suponer también que el valor inicial de la asimetría bariónica fue enorme; los cálculos arrojan un número, 10^{69}, que por su inmensidad resulta aún menos natural que la pequeña asimetría bariónica actual. Así pues, la inflación hace muy probable que la asimetría bariónica requiera una explicación dinámica.



EMMA SKURNICK/*American Scientist*

3. **LOS BARIONES** son estados ligados de los quarks. Se dice que los quarks tienen "sabores": arriba y abajo, encantado y extraño, cima y fondo. El neutrón consta de dos quarks abajo y un quark arriba. El antineutrón está formado por los correspondientes antiquarks (su símbolo coincide con el de los quarks respectivos, pero con una barra encima).

Las leyes de Sajarov

Mientras Sajarov trabajaba en el problema de la bariogénesis, otro científico ruso, Vadim Kuzmin, identificó las condiciones necesarias. Iba tres años por detrás de Sajarov al proponer una teoría en 1970, pero aún siete por delante del resto del mundo; hasta 1977 no empezaron los teóricos a tomarse estas ideas en serio. En sus trabajos, Sajarov y Kuzmin establecieron tres condiciones que debían satisfacerse para que la bariogénesis tuviera lugar.

En primer lugar, el número bariónico no se debía conservar. Es decir, habrá interacciones que cambien el número del bariones del universo.

En segundo, deben violarse dos simetrías que relacionan partículas con antipartículas.

En tercero, debe haber una pérdida de equilibrio termodinámico.

El segundo y el tercer requisito son bastante técnicos. Siempre entiendo mejor los conceptos abstractos cuando se aplican a un ejemplo; expondré, pues, la razón de la necesidad de estas leyes mediante uno de los primeros esquemas de bariogénesis que se formularon, las Teorías de Gran Unificación (o GUT, de "Grand Unified Theories").

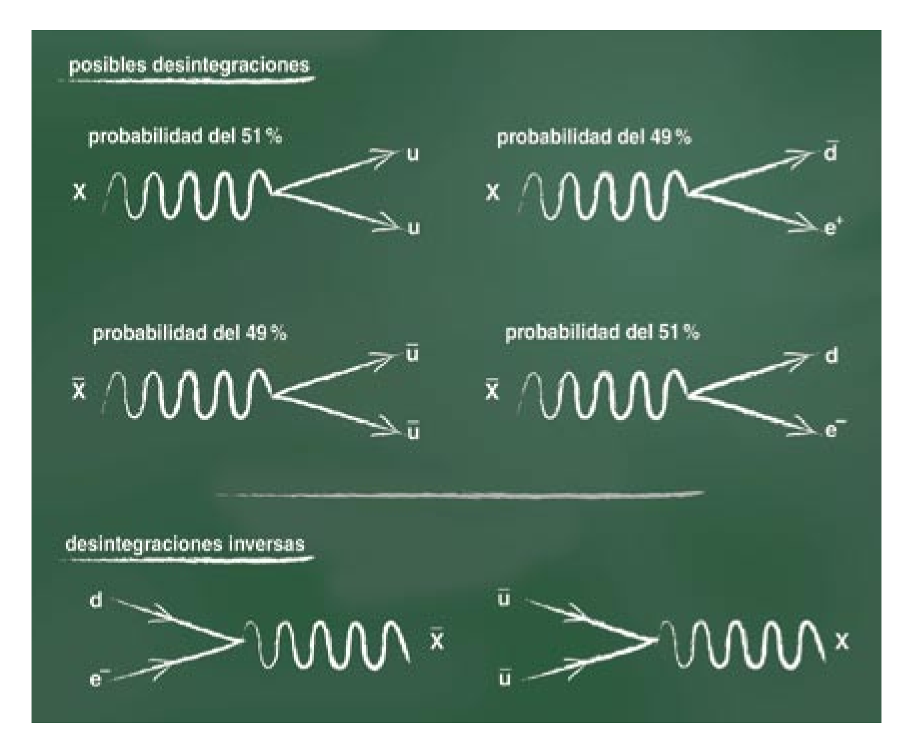
La idea básica en que se funda la bariogénesis en las GUT es bastante simple: unas partículas muy pesadas, los bosones X, dejan al desintegrarse a altas temperaturas más quarks que antiquarks. Por ejemplo, supongamos que la partícula X sigue un cierto proceso de desintegración el 51 % de las veces, mientras que su antipartícula, $\overline{X}$, sigue el camino de desintegración paralela el 49 % de las veces. Esta asimetría provocaría rápidamente un exceso de bariones.

En la figura 4 se muestra una situación verosímil. Los quarks que constituyen los bariones ordinarios, el protón y el neutrón, aparecen en dos "sabores": "arriba" y "abajo"; al haber tres en cada barión, el número bariónico de un quark es 1/3. El positrón es la antipartícula del electrón. Supongamos que el bosón X se desintegra el 51 % de las veces en dos quarks arriba y el 49 % en un antiquark abajo y un positrón. En promedio, cada desintegración de un X producirá $(2/3 \times 0{,}51) - (1/3 \times 0{,}49) = 0{,}177$ bariones (al calcular el número bariónico las antipartículas se restan).

Su antipartícula $\overline{X}$ se desintegra el 49 % de las veces en dos antiquarks arriba y el 51 % en un quark abajo y un electrón. Esto da $(-2/3 \times 0{,}49) + (1/3 \times 0{,}51) = -0{,}157$ bariones. Cien desintegraciones de cada partícula aportan una ganancia neta de dos bariones, muy superior a lo que se necesita para generar la pequeña asimetría bariónica que observamos.

¿Por qué se requieren las leyes de Sajarov para que valga este esquema?

Violación del número bariónico. Si X sólo pudiera desintegrarse en dos quarks, diríamos que tiene un número bariónico de 2/3. Si X sólo pudiera desintegrarse en un único antiquark, le asignaríamos un número de –1/3 (ya que los antiquarks tienen números bariónicos negativos). Que ambos canales sean viables significa que no se puede dar un número bariónico inequívoco a la partícula. Sólo sería posible si todas las interaccio-



4. LAS PARTICULAS SUBATOMICAS dejan, al desintegrarse, partículas de materia y energía. Puede surgir una asimetría si la probabilidad de un cierto canal de desintegración es distinta para una partícula y para su antipartícula. Tomemos como ejemplo una partícula hipotética, el bosón X. Supongamos que es un poco más probable que X se desintegre en dos quarks arriba que en un antiquark abajo y un positrón, y que las probabilidades de los caminos de desintegración paralelos de su antipartícula $\bar{X}$ están invertidas: es un poco más probable que $\bar{X}$ se desintegre en un quark abajo y un electrón que no en dos antiquarks arriba. Cien desintegraciones de cada partícula darían una ganancia neta de dos bariones. El panel inferior ilustra las desintegraciones inversas, la fusión de partículas para crear nuevos bosones X. A temperaturas elevadas en el universo primitivo, las desintegraciones inversas serían tan rápidas como las desintegraciones, lo que arruinaría la asimetría.

EMMA SKURNICK/*American Scientist*

nes respetaran el número bariónico, tal como sucede en la física usual.

Violación de las simetrías de conjugación de carga. Ya he mencionado antes que ciertas propiedades de las partículas no deben ser perfectamente simétricas; dos de esas propiedades cuasisimétricas se conocen con las letras C y CP. Con C se simboliza la conjugación de carga, la operación que cambia una partícula en una antipartícula; la letra P nos remite a la paridad, la operación que cambia el signo de las coordenadas espaciales de un sistema y crea una imagen especular. Algunas teorías físicas que rigen las propiedades microscópicas permanecen invariables cuando se cambian las partículas por las antipartículas. Pero en el ejemplo anterior he dispuesto una asimetría al dejar que un barión y su antibarión prefirieran distintos modos de desintegrarse.

Pérdida de equilibrio térmico. Un buen ejemplo de equilibrio térmico es el agua hirviendo dentro de una olla a presión cerrada. Las moléculas de agua pasan de la fase líquida a la gaseosa sin cesar. Hay equilibrio porque, una vez se ha alcanzado una temperatura constante, el ritmo a que las moléculas hacen esta transición es exactamente igual al ritmo del proceso inverso, de manera que la cantidad total de líquido y vapor permanece constante. Se rompe este equilibrio térmico si se levanta la tapa. El vapor escapa y el ritmo de transformación de líquido a gas se vuelve mayor que el ritmo inverso. Si se sigue suministrando calor, se evaporará todo el líquido.

Hay una situación en el universo primitivo análoga a levantar la tapa de la olla. En el equilibrio, los bariones se están desintegrando, pero acontece también el proceso inverso, en el cual los quarks se unen para formar bariones; ambos procesos suceden al mismo ritmo. Las desintegraciones no pueden producir una asimetría bariónica, ya que las desintegraciones inversas la van deshaciendo. Sin embargo, a medida que el universo se expande, la temperatura decrece, como si se destapara la olla a presión. A determinada temperatura un par de quarks ya no tendrá la suficiente energía como para producir una partícula pesada. Se generará así una asimetría bariónica.

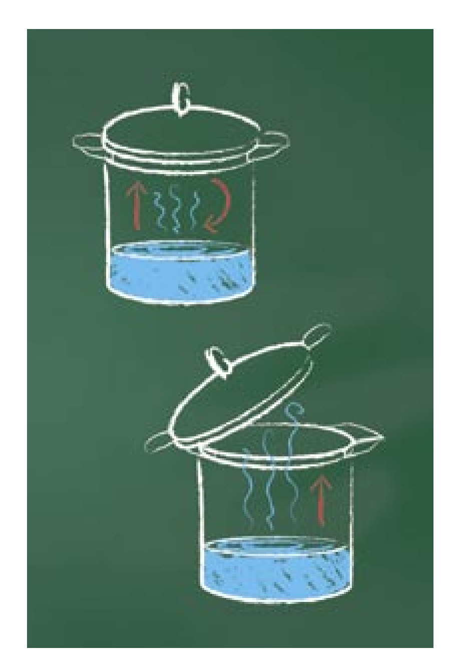
Violación de CP

Sajarov indicó que, para la bariogénesis, no se debía violar un tipo de simetría, sino dos. De ellos uno requiere cierta explicación. La violación de CP combina dos transformaciones. C cambia partículas por antipartículas. Su vinculación con la paridad resulta menos obvia.

Las posiciones de los objetos en el espacio se describen mediante tres ejes, x, y y z. Cuando se mira en un espejo, se ve una transformación de paridad en el eje perpendicular a la superficie del espejo. Si es el eje y y el objeto real está en la posición x, y y z, la posición especular del objeto se encontrará en la posición x, $-y$ y z. Un reflector de esquina, consistente en tres espejos formando un vértice de cubo, muestra la completa transformación de paridad ($-x$, $-y$ y $-z$) en las tres direcciones.

La paridad es relevante para la bariogénesis debido a que los quarks no sólo tienen sabores; también pueden ser levógiros o dextrógiros. Una manera sencilla de imaginarlo: si los dedos de la mano derecha se curvan en torno a una partícula imaginaria que se mueve en la dirección del dedo pulgar estirado, una partícula dextrógira girará en la dirección en la que apuntan los dedos. Una partícula levógira girará en sentido opuesto.

Si la simetría CP se hubiese satisfecho durante toda la historia del universo, la única asimetría de partículas que crearíamos sería entre quarks levógiros y antiquarks dextrógiros, más una asimetría igual y opuesta entre quarks dextrógiros y antiquarks levógiros. Por desgracia, la asimetría total entre quarks y antiquarks sería aún nula.



5. ANDREI SAJAROV MOSTRO QUE UNO DE LOS REQUISITOS para la creación de materia es la pérdida del equilibrio térmico. Si las consideraciones térmicas se añaden a la situación hipotética ilustrada en la figura 4, puede surgir una asimetría bariónica. En equilibrio térmico está el interior de una olla cerrada: el líquido entra y sale de la fase gaseosa a velocidades iguales. La expansión del universo equivale a quitar la tapa de la olla. A medida que el universo se enfría, alcanza una temperatura en la que dos quarks carecen ya de suficiente energía para producir un bosón pesado X.

La búsqueda de violaciones de la simetría CP ha sido una pasión de los físicos de partículas desde los años cincuenta. Antes se creía que CP era una simetría exacta. Se vio lo erróneo de esta creencia cuando los experimentos de 1964 de James Christenson, James Cronin, Val Fitch y Rene Turlay, del Laboratorio Nacional de Brookhaven, encontraron un proceso que violaba la simetría CP, aunque sólo un poco. Dos partículas supuestamente distintas resultaron ser una misma, el kaón. Se demostró que el kaón neutro se desintegraba en estados con diferentes valores de CP.

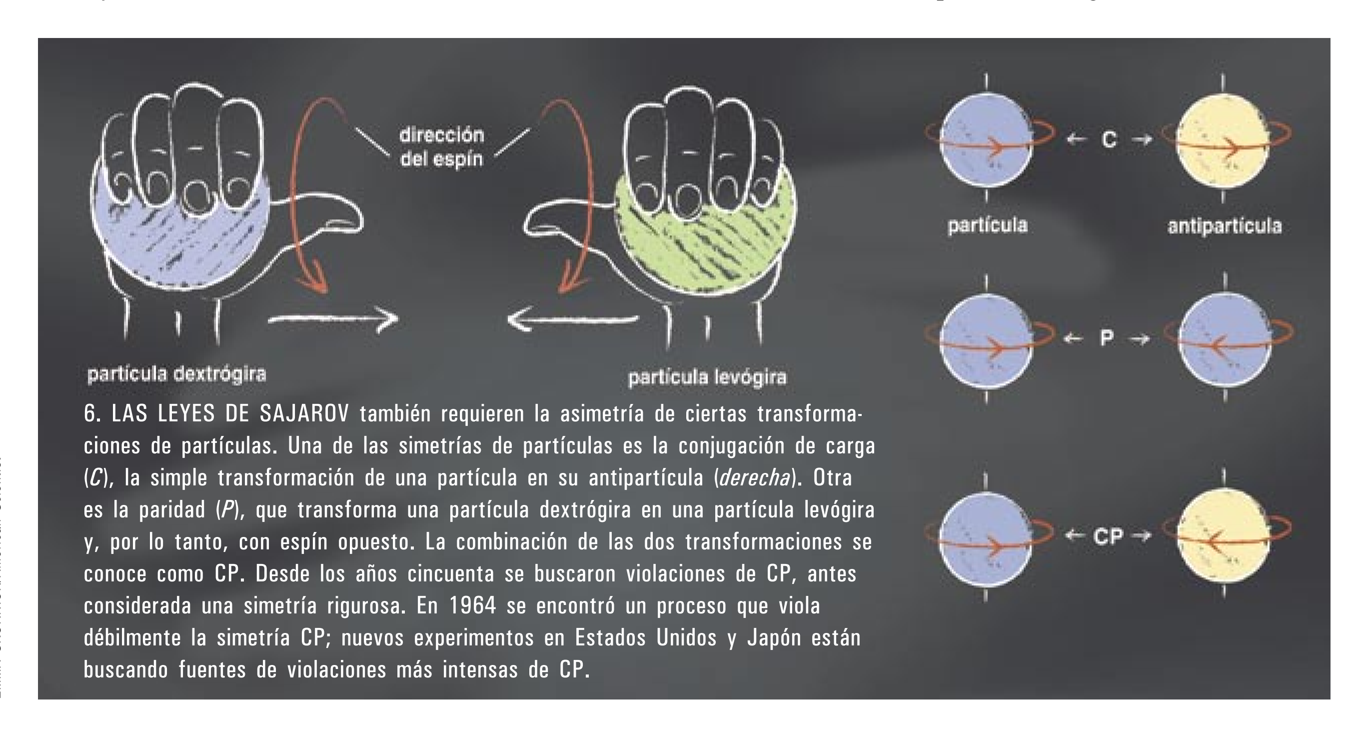
Sin embargo, la violación de CP en las desintegraciones de los kaones es tan débil, que ni siquiera puede crear una asimetría bariónica tan pequeña como la que se observa. Se espera que se descubran fuentes mayores de violación de CP. El experimento BaBar, del Centro del Colisionador Lineal de Stanford, y un experimento japonés análogo, Belle, están ahora buscando violaciones de CP en las interacciones de ciertos quarks pesados.

Burbujas y masa

Con el descubrimiento experimental de la violación de CP, Sajarov tuvo una realidad física que le inspirase ideas acerca del origen de la materia. No sabía que otro de sus criterios de violación de simetrías, la necesidad de violación del número bariónico, se cumpliría en el modelo estándar de la física de partículas por medio de cierto proceso, el esfalerón (del griego "caer").

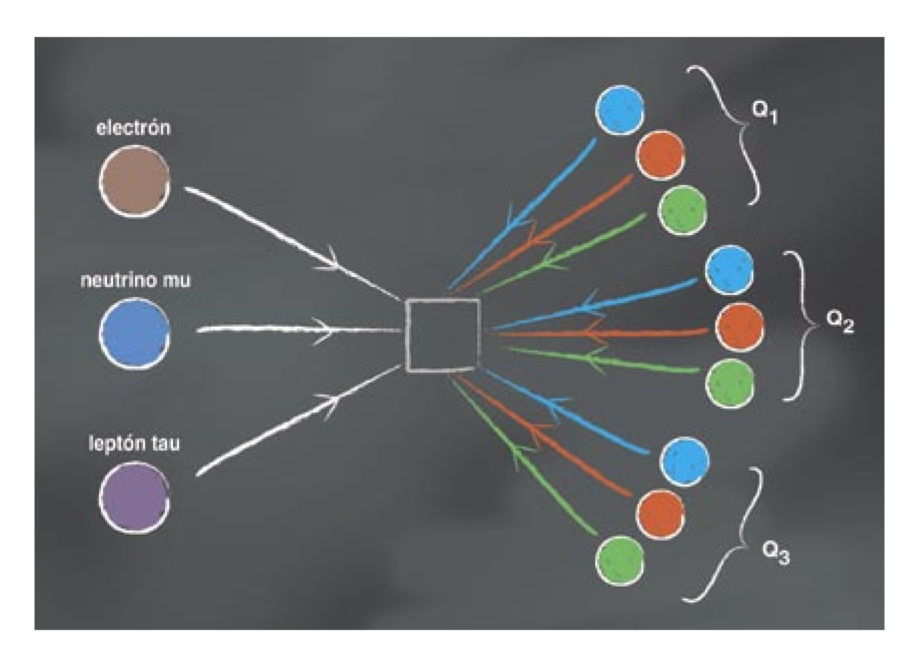
El modelo estándar unifica el electromagnetismo y la fuerza nuclear débil en unas interacciones electrodébiles, con las correspondientes partículas. El esfalerón es cierta interacción concreta predicha por el modelo estándar. Involucra nueve quarks y tres partículas ligeras de la clase de los leptones, que se pueden contar de manera parecida a como se cuentan los bariones. Descubierto como resultado matemático en 1984 por Frans Klinkhamer y Nicholas Manton (ahora en las Universidades de Karlsruhe y Cambridge, respectivamente), el esfalerón no se ha confirmado experimentalmente, aunque ningún teórico duda de su validez.

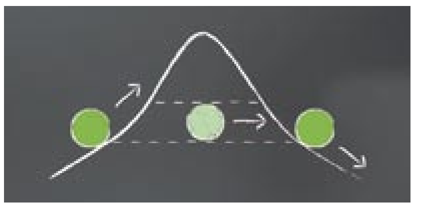
Para que tenga lugar la violación bariónica, se juntan tres quarks de cada una de las "generaciones" del modelo estándar —la de los dos quarks ligeros (*arriba* y *abajo*) y otras dos de quarks más pesados, descubiertos en los experimentos a altas energías de los aceleradores— con un leptón de cada generación de leptones correspondiente. A baja temperatura, este proceso requiere que se produzca un efecto túnel cuántico, una transición del sistema microscópico que normalmente requeriría un aporte de energía.



6. LAS LEYES DE SAJAROV también requieren la asimetría de ciertas transformaciones de partículas. Una de las simetrías de partículas es la conjugación de carga (*C*), la simple transformación de una partícula en su antipartícula (*derecha*). Otra es la paridad (*P*), que transforma una partícula dextrógira en una partícula levógira y, por lo tanto, con espín opuesto. La combinación de las dos transformaciones se conoce como CP. Desde los años cincuenta se buscaron violaciones de CP, antes considerada una simetría rigurosa. En 1964 se encontró un proceso que viola débilmente la simetría CP; nuevos experimentos en Estados Unidos y Japón están buscando fuentes de violaciones más intensas de CP.







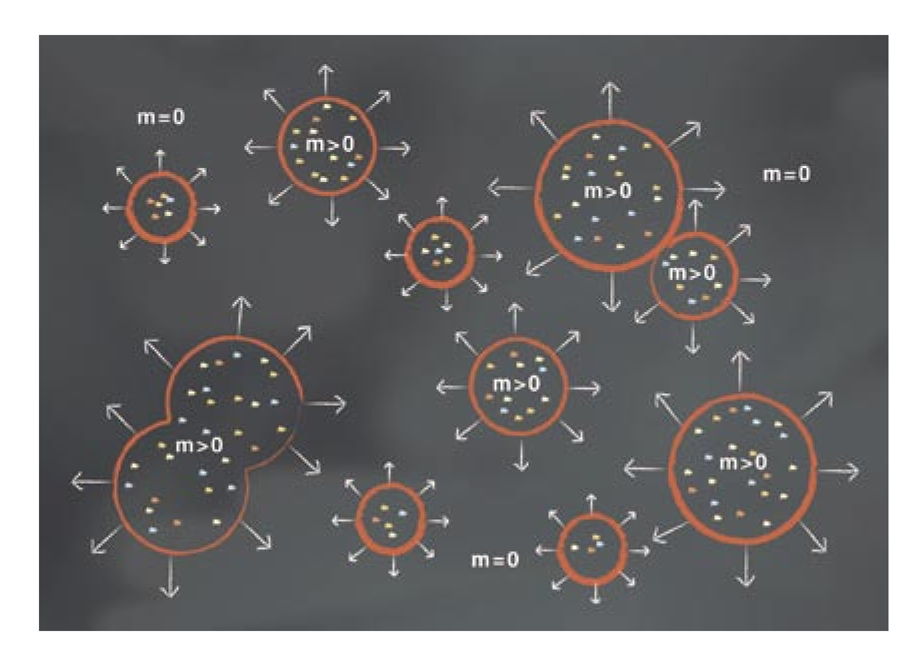
7. LA VIOLACION DEL NUMERO BARIONICO quizá se deba a una determinada interacción entre quarks y leptones, dos grandes clases de partículas elementales. En ella intervendrían quarks de las tres generaciones del modelo estándar de las partículas elementales y un leptón de cada una de las correspondientes generaciones de leptones (*izquierda*). A bajas temperaturas, esa interacción requiere un "túnel cuántico", donde el sistema acomete una transición que, en condiciones normales, demandaría una aportación de energía (*arriba*). Figuradamente hablando, el sistema abre un "túnel" a través de una barrera de energía. A temperaturas más altas, la transición puede sobrepasar la barrera: se realiza una interacción de esfalerón.

La probabilidad de ese efecto túnel es tan pequeña, que no se esperaría ver que ocurriese en el laboratorio o en el universo observable. ¿De dónde le viene su importancia, pues? Dieron la respuesta Kuzmin, Valery Rubakov y Mikhail Shaposhnikov en 1985: a las altas temperaturas del universo primitivo, la barrera energética que evita la violación bariónica podría superarse con energía térmica. Es decir: a altas temperaturas sí ocurre la interacción esfaleróníca. Aún más importante es la disminución, e incluso anulación, de la barrera energética a temperaturas suficientemente elevadas.

La aceptación general de la validez del esfalerón por parte de los teóricos significa que dos de los tres criterios de Sajarov —la violación de la simetría CP y la violación del número bariónico— se cumplen en la naturaleza. ¿Existe un proceso que satisfaga la tercera condición, la pérdida del equilibrio térmico? Si así fuera, comprenderíamos el tamaño de la asimetría bariónica a partir de principios conocidos, sin tener que establecer nuevas hipótesis, y dispondríamos de condiciones restrictivas para cualesquiera nuevas ideas físicas que predijesen un valor distinto de la asimetría bariónica.

Antes el agua hirviente ha servido de ejemplo de pérdida del equilibrio térmico. Hay un proceso con ese mismo efecto en la teoría electrodébil a altas temperaturas, pero opuesto a lo que sucede cuando el agua hierve. En el universo primitivo se formaron burbujas a medida que el universo se enfriaba. La fase existente fuera de las burbujas no era usual; en esa fase las partículas carecían de la masa que tienen de ordinario. Sólo dentro de las burbujas las partículas recuperaban sus masas y generaban la física que conocemos. Las burbujas se expandieron y, finalmente, excluyeron del universo toda la fase exótica sin masa. Un proceso de este tipo constituye una transición de fase de primer orden.

¿Qué relación guarda ello con la bariogénesis? En esta teoría los esfa-

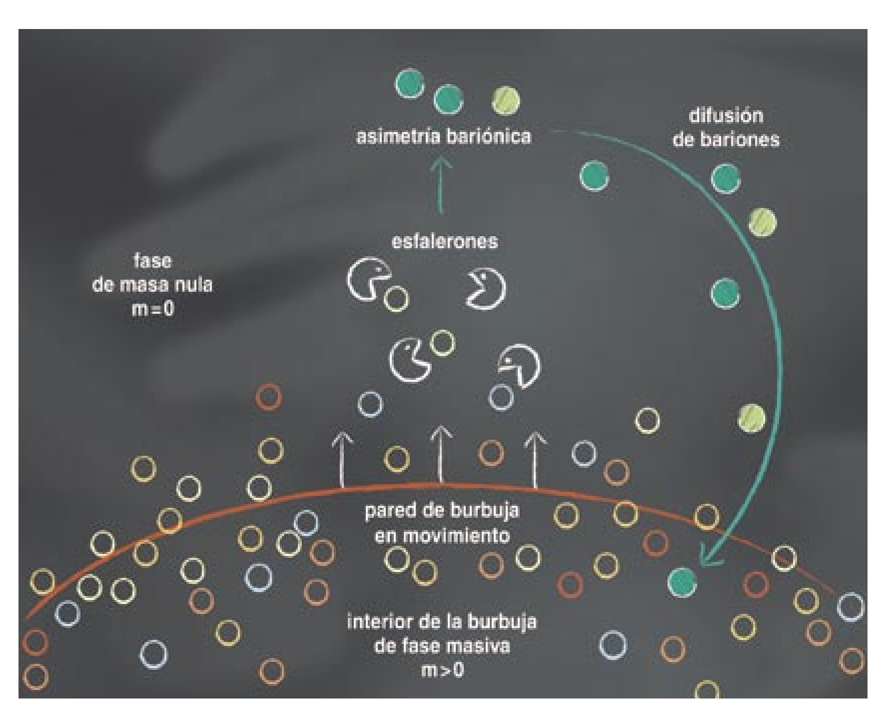


8. SEGUN LA TEORIA ELECTRODEBIL, se formaron burbujas mientras las altas temperaturas del universo primitivo se enfriaban. En su interior, las partículas tenían masa y la física era la ordinaria; la fase de fuera de las burbujas carecía de masa. Las burbujas se expandieron hasta que por fin eliminaron la fase exótica sin masa mediante una "transición de fase de primer orden".

EMMA SKURNICK/*American Scientist*

lerones, la fuente de la violación del número bariónico, son mucho más débiles dentro de las burbujas (en el reino de la física ordinaria) que a extramuros de las mismas. Fuera no existe la “barrera” de energía que en condiciones normales los esfalerones han de pasar por efecto túnel; suceden sin ninguna barrera que superar. Por tanto, la violación bariónica es fuerte fuera de las burbujas y mucho más débil dentro. Finalmente, como se recordará, a bajas temperaturas los esfalerones se debilitan tanto, que resultan despreciables.

En la teoría electrodébil, al expandirse las burbujas y llenar todo el espacio, cada partícula del universo acaba por atravesar la pared de una de ellas hacia su interior. Andrew Cohen, de la Universidad de Boston, junto con David B. Kaplan y Ann Nelson, ahora en la de Washington, percibieron en 1990 que así se crea un mecanismo para la bariogénesis. Cuando los quarks encuentran la pared tienen cierta probabilidad de penetrar en el interior de la burbuja o de rebotar hacia el exterior. La violación de CP permite que esta probabilidad difiera para quarks y antiquarks, y para quarks levógiros y dextrógiros. Supongamos que se crea una asimetría de los antiquarks levógiros en la fase sin masa a extramuros de las burbujas. Los esfalerones intentarán borrar esta asimetría; mas, al hacerlo, cambiarán el número bariónico total, generando una asimetría bariónica. Finalmente, esos bariones caerán dentro de las burbujas, que llenarán el universo.



9. EN LA SITUACION DESCRITA en la figura 8, los procesos esfaleróniсos son mucho más débiles dentro de las burbujas que fuera, donde no tienen que atravesar por efecto túnel una barrera de energía. Esa diferencia proporciona un mecanismo para la bariogénesis. Podría establecerse una asimetría de los antiquarks levógiros en la fase sin masa. Los esfalerones intentarían borrar la asimetría, pero así cambiarían el número bariónico total, creando una asimetría bariónica. Los bariones se difundirían dentro de las burbujas que llenan el universo. Por desgracia, en el marco del modelo estándar de la física de partículas las burbujas sólo se forman si ciertas partículas son más bien ligeras: recientes experimentos realizados en aceleradores de partículas han probado que no son tan livianas como se necesitaría.

EMMA SKURNICK/*American Scientist*

Los esfalerones proceden con lentitud mayor dentro de la burbuja que fuera de ella. Como los ritmos no son iguales, los esfalerones quedan fuera del equilibrio en el interior de la burbuja. Tan decisivo resulta ello, que si no ocurriera así los esfalerones del interior de la burbuja destruirían la asimetría bariónica creada por los que están fuera.

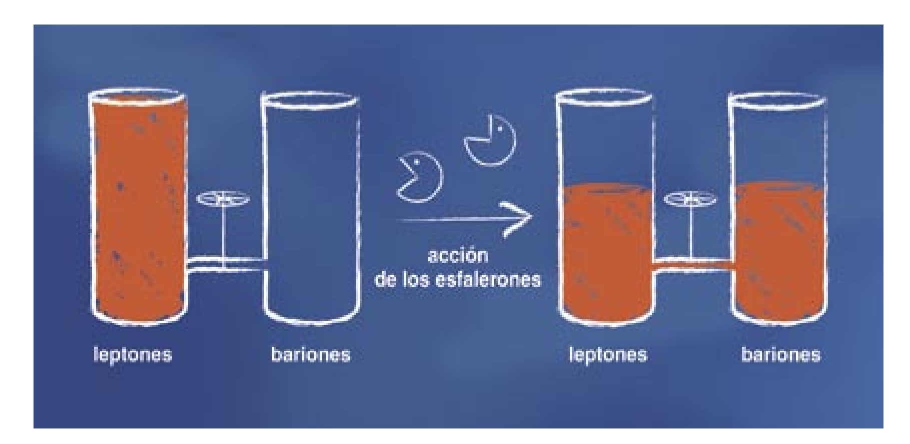
En este repaso crítico he esbozado la trayectoria que seguí con mi colaborador Kimmon Kainulainen, de la Universidad de Jyväskylä, de Finlandia, en nuestra reciente búsqueda fallida de una teoría correcta de la bariogénesis. Llegamos a una manera de comprobar la teoría, a una pregunta cuantitativa: los esfalerones de dentro de las burbujas de nuestro universo ¿son tan lentos como para producir la asimetría bariónica observada? Ello depende del tamaño de la barrera energética que deban superar; o, en el lenguaje de los físicos, de la intensidad de la transición de fase. Si la transición es demasiado débil, la asimetría bariónica se reduce sobremanera.

Shaposhnikov y otros encontraron que la intensidad de la transición de fase depende, en el modelo estándar, de la ligereza de ciertas partículas: el bosón de Higgs y el quark “cima”. Recientes resultados de experimentos con aceleradores han demostrado que son bastante pesadas; en consecuencia, se debe admitir que la transición de fase resulta demasiado débil para explicar la bariogénesis, a no ser que se añadan nuevos principios físicos hipotéticos. No todas las condiciones de Sajarov pueden satisfacerse dentro del modelo estándar.

Supersimetría

¿Qué nueva física nos ayudaría? Una de las ideas mejor motivadas, la supersimetría, o SUSY, nació en 1970 en el Instituto Lebedev; la trajeron al mundo Yuri Golfand y sus colaboradores. La supersimetría extiende las simetrías que caracterizan las familias de partículas del modelo estándar y establece que, para cada clase conocida de partícula que tenga un cuanto de espín, es el caso de los quarks y los leptones, existe una supercompañera de espín nulo. No se han descubierto aún tales partículas; por tanto, se cree que tienen masas grandes, más allá de nuestra capacidad de detección. Pero SUSY logra explicar —en realidad se inventó para explicarlo— otro de los misteriosos números pequeños del modelo estándar, la masa de la partícula Higgs.

Las nuevas partículas predichas por la supersimetría aumentan la intensidad de la transición de fase



10. LA LEPTOGENESIS, la creación de las partículas ligeras conocidas como leptones, es una de las vías posibles para la creación de la materia. Si hay una asimetría entre ciertos leptones y sus antipartículas, los procesos esfalerónicos pueden convertir la asimetría leptónica en una asimetría bariónica. La interacción esfalerónica vendría a ser como abrir una válvula que igualase la asimetría leptónica y la bariónica. Los estudios relativos a los neutrinos, los leptones más fantasmagóricos, han aportado pruebas de que tienen masa, lo que ha alimentando el interés en la leptogénesis en cuanto clave de la bariogénesis. En tal caso, los objetos de mayor masa del universo se habrían originado gracias a las partículas más etéreas que se conocen.

EMMA SKURNICK/*American Scientist*

electrodébil al elevar la barrera del esfalerón y salvar de los estragos de los esfalerones no suprimidos a los bariones que se hallan en las burbujas. SUSY proporciona nuevas fuentes de violación de CP, más intensas que los procesos demasiado débiles del modelo estándar. Pero esta propiedad se ha convertido en el talón de Aquiles del modelo. Los mismos procesos que violan la simetría CP crean momentos dipolares eléctricos de quarks, electrones, neutrones y núcleos atómicos bastante grandes. Esos momentos deben cumplir las restricciones impuestas por los experimentos, que sólo miden sus límites superiores. Por un tiempo, se pudo imaginar que las fuentes supersimétricas de violación de CP se conjugasen de modo que crearan una asimetría bariónica grande y al mismo tiempo diesen momentos dipolares eléctricos pequeños. Con el tiempo, la acumulación de datos experimentales ha ido, sin embargo, quitando verosimilitud a estas expectativas.

La mayoría de los expertos considera hoy que la versión mínima de la bariogénesis electrodébil basada en la supersimetría está muerta o, por lo menos, es muy improbable. Cabría resucitarla complicando el modelo mínimo con partículas e interacciones extra. Quizá se descubran tales añadidos cuando entre en funcionamiento en 2007 el que será el mayor acelerador de partículas del mundo, el Gran Colisionador de Hadrones del Laboratorio Europeo de Física de Partículas, el CERN, en Ginebra. Hasta entonces, esas versiones adornadas de la teoría permanecerán en el reino de la especulación.

Leptogénesis

Tras la agonía de la bariogénesis electrodébil, otra idea ha venido a ocupar su puesto: la bariogénesis vía *leptogénesis*. Aunque la teoría electrodébil no puede satisfacer fácilmente con bariones las tres condiciones de Sajarov, no resulta difícil adaptar el argumento para crear una asimetría de otras clases de partículas, en concreto la de los leptones, que incluye electrones, muones, partículas tau y neutrinos. Se puede establecer una asimetría entre neutrinos y antineutrinos. Pero estamos constituidos por bariones, no por fantasmagóricos neutrinos; ¿de qué le sirve esa asimetría neutrínica a la bariogénesis?

De nuevo el proceso del esfalerón ofrece un instrumento para pensar acerca de la asimetría. Recordemos que en los esfalerones participan tanto leptones como quarks. Por esta razón, es casi imposible crear una asimetría leptónica a altas temperaturas sin que se convierta, al menos parcialmente, en una asimetría bariónica. Poner en marcha las interacciones esfalerónicas es abrir una válvula que iguala las asimetrías leptónica y bariónica.

La leptogénesis se apoya en la desintegración de los neutrinos pesados, partículas hipotéticas cuya existencia se requiere para explicar por qué los ligeros neutrinos del modelo estándar apenas tienen masa. La carencia absoluta de carga eléctrica y casi completa de masa de los neutrinos del modelo estándar los vuelve fantasmagóricos. Aunque en el universo abundan casi tanto como los fotones, muy rara vez interaccionan con otras partículas. Ha costado estudiar si tienen masa y hasta qué punto es pequeña. Se han usado toda clase de ingeniosas técnicas para observar los neutrinos del Sol, de las interacciones de los rayos cósmicos con la atmósfera de la Tierra y de los reactores nucleares. Cada vez hay más pruebas de que los neutrinos tienen masas no nulas. Este hallazgo alimenta el interés en la leptogénesis en cuanto clave de la bariogénesis.

Por desgracia, no está claro que la leptogénesis pueda comprobarse directamente. Que las masas conocidas de los neutrinos cumplan los requisitos de la leptogénesis da esperanzas, pero dista de ser una prueba. Podemos esperar que la futura observación de raras desintegraciones —la del pesado muon, por ejemplo, en un electrón y un fotón— proporcionará más pruebas circunstanciales de la leptogénesis. Quizá sólo podamos esperar signos que apunten hacia la leptogénesis y nunca tengamos una verdadera prueba.

De cuerdas e ideas postergadas

En nuestro empeño por entender la naturaleza del universo, los teóricos a menudo debemos admitir que nos hemos introducido en un callejón quizá sin salida, que hemos dado con una pregunta a la que nunca podremos contestar de manera satisfactoria. La esperanza de muchos cosmólogos de resolver tales cuestiones radica en la búsqueda de una teoría fundamental que haga predicciones universales, de las que se deducirían sin ambigüedades las respuestas de las cuestiones menores.

Los desarrollos recientes en una teoría fundamental, la teoría de cuerdas, nos hace sospechar que nuestras preguntas sobre números pequeños cuenten con una multitud de respuestas, no con una sola, inequívoca, o que haya una plétora de teorías que se apliquen en regiones diferentes del universo, regiones que, a efectos prácticos, se podrían considerar universos separados, ya que no se comunicarían entre sí. No tendríamos manera, *a priori*, de predecir qué fase de la teoría se aplica en la región que nos ha tocado.

Sin embargo, sólo algunas de las fases disponibles en la teoría de cuerdas son compatibles con nuestra existencia, ya que otras predicen universos con condiciones hostiles al desarrollo de vida. Algunos querrían restringir la atención a las fases de la teoría de cuerdas compatibles con nuestra propia existencia. Muchos consideran este "principio antrópico" una renuncia a hacer ciencia de verdad. Su reaparición en la teoría de cuerdas ha polarizado a los físicos teóricos.

La existencia humana restringe muchos de los parámetros de las leyes que gobiernan la física que conocemos. Estas leyes deben definirse de manera que se ajusten a lo que observamos, un universo que contiene vida. La asimetría bariónica, ¿es antrópicamente dependiente? Es decir, ¿qué sacaríamos de buscar sólo valores compatibles con nuestra propia existencia? Se ha argumentado que valores comprendidos entre 10^{-4} y 10^{-11} cumplen esa condición: permitirían que hubiese estrellas y galaxias (lo que no sería posible si escasease demasiado la materia), e impedirían que la producción de helio superara la de hidrógeno (como sucedería si la asimetría fuese demasiado grande), en cuyo caso no habría agua y las estrellas se consumirían antes de que la temperatura del universo llegara a ser la adecuada para la vida. Debido a que estas limitaciones dejan un intervalo bastante grande de posibles valores, los argumentos antrópicos no afectan demasiado a la teoría, así que podemos dejar de lado el debate antrópico y seguir buscando una explicación dinámica y determinista del origen de la materia.

Tendría su gracia que la teoría más en boga hoy, la leptogénesis, fuese cierta. Significaría que los objetos de mayor masa del universo tendrían su origen en las partículas más etéreas que conocemos, los neutrinos. Por mi parte, aún pongo esperanzas en la idea postergada, la bariogénesis electrodébil supersimétrica. Puede que todavía haya atajos que no se hayan visto y maneras de satisfacer las limitaciones experimentales impuestas a la teoría, y quizá se pueda comprobarla con más detalle cuando el Gran Colisionador de Hadrones empiece a funcionar.

El autor

James M. Cline es profesor de física en la Universidad McGill, en Montreal. Ha trabajado en varias áreas de la física de las partículas elementales, incluyendo sus implicaciones para la cosmología primitiva así como la presente expansión acelerada del universo.

Bibliografía complementaria

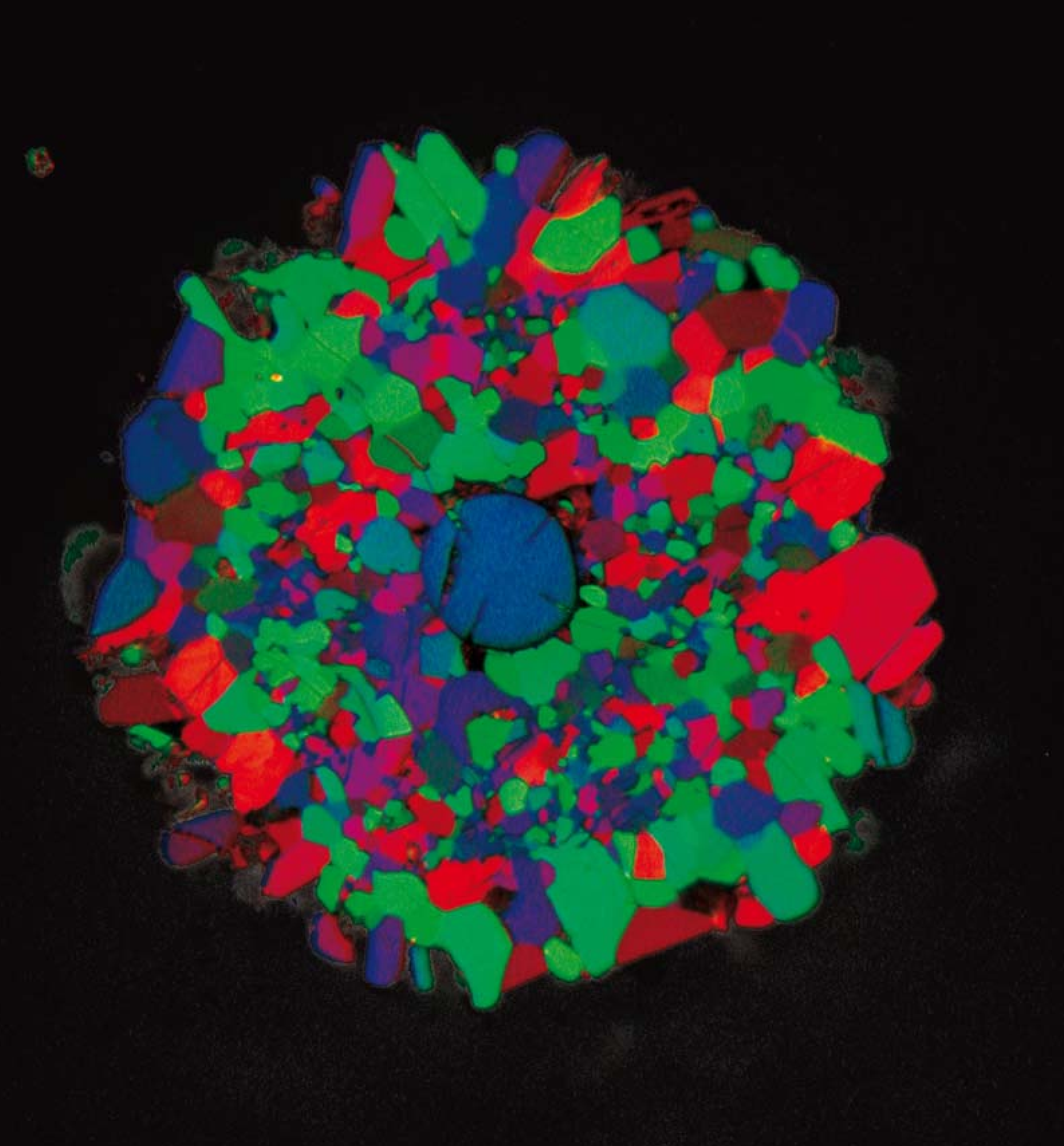
VIOLATION OF CP INVARIANCE, C ASYMMETRY, AND BARYON ASYMMETRY OF THE UNIVERSE. A. D. Sakharov en *JETP Letters*, vol. 5, págs. 24-27; 1967.

CP-NONINVARIANCE AND BARYON ASYMMETRY OF THE UNIVERSE. V. A. Kuzmin en *Journal of Experimental and Theoretical Physics*, vol. 12, págs. 228-230; 1970.

ON THE ANOMALOUS ELECTROWEAK BARYON NUMBER NON-CONSERVATION IN THE EARLY UNIVERSE. V. A. Kuzmin, V. A. Rubakov y M. E. Shaposhnikov en *Physics Letters B*, vol. 155, págs. 36-42; 1985.

SUPERSYMMETRIC ELECTROWEAK BARYOGENESIS IN THE WKB APPROXIMATION. J. M. Cline, M. Joyce y K. Kainulainen en *Physics Letters B*, vol. 417, págs. 79-96; 1998.

ORIGIN OF THE MATTER-ANTIMATTER ASYMMETRY. M. Dine y A. Kusenko en *Reviews of Modern Physics*, vol. 76, págs. 1-30; 2004.



El diboruro de magnesio, superconductor a alta temperatura

El diboruro de magnesio desafía los principios que hasta hace muy poco se suponía que cumplían los superconductores óptimos. Con una temperatura crítica cercana a los 40 kelvin, promete una gran variedad de aplicaciones

Paul C. Canfield y Sergey L. Bud'ko

HAL SAILSBURY *Materials Preparation Center, Ames Laboratory*

SECCION TRANSVERSAL PULIDA de un segmento de cable de diboruro de magnesio de 0,14 milímetros de diámetro. El cable, compacto, está formado por pequeños granos carentes de orientación que reflejan la luz de modo dispar, dando lugar a varios colores. Este tipo de cables resultan útiles para caracterizar la superconductividad del material.

Imagine que en su jardín encuentra una veta de oro o un pozo de petróleo. Una sensación similar de incrédula excitación se extendió sobre la comunidad de físicos del estado sólido durante las primeras semanas de 2001, cuando se anunció que el diboruro de magnesio (MgB_2) se torna superconductor —esto es, conduce la electricidad sin resistencia eléctrica— a temperaturas cercanas a los 40 grados kelvin.

Este compuesto simple se conocía desde el ecuador del siglo pasado, pero había permanecido en los estantes de los laboratorios durante decenios sin que nadie sospechara su valiosísimo potencial. Aunque 40 K (o –233 ºC) pueda parecer una temperatura baja, casi doblaba el récord de los compuestos metálicos (alrededor de 23 K para las aleaciones de niobio, utilizadas extensamente en investigación e industria). En el MgB_2, la temperatura de transición, o temperatura crítica T_c, se alcanza mediante técnicas harto más baratas que las que requieren las aleaciones de niobio. Entre las posibles aplicaciones del nuevo superconductor se incluyen bobinas superconductoras (imanes) y cables del tendido eléctrico.

A diferencia de los superconductores de altas temperaturas (los óxidos de cobre se tornan superconductores a temperaturas de hasta 130 K), el MgB_2 parece ser un superconductor tradicional, aunque una nueva variante. En la búsqueda de superconductores con temperaturas de transición cada vez superiores, los físicos habían deducido algunas leyes simples que les permitían predecir qué clase de combinaciones químicas podían dar buen resultado. Además, se sospechaba que 23 K era una temperatura de transición muy próxima a la máxima posible en un superconductor normal. Para sorpresa de todos, el MgB_2 desafió dichas leyes e hizo volar por los aires la barrera conceptual que impedía conseguir temperaturas superiores.

Las investigaciones sobre el MgB_2 avanzaron a una velocidad asombrosa. Jun Akimitsu, de la Universidad Aoyama Gakuin en Tokio, anunció su descubrimiento en un congreso a mediados de enero de 2001. Tan sólo dos meses después, se presentaban dos breves intervenciones sobre el tema en la reunión anual de la Sociedad Americana de Física y se habían añadido más de 70 artículos en el sitio de la Red "arxiv.org".

Esta explosión de actividad obedecía a varias razones. Por un lado, una vez se conoce la receta, es relativamente fácil sintetizar MgB_2 con un buen nivel de pureza. Por otro, en 2001, gracias a Internet, los físicos de la materia condensada nunca habían estado mejor comunicados entre sí. Estos dos ingredientes, combinados con la promesa de un nuevo y simple superconductor de temperatura crítica elevada, propiciaron una mezcla intelectual explosiva.

La confirmación del descubrimiento

Al principio, las noticias del anuncio de Akimitsu se extendieron tan sólo de boca en boca y por correo electrónico. No había disponibles ninguna publicación científica ni borrador electrónico. Cuando la noticia llegó a nuestro grupo, unos pocos días después de la reunión, varias preguntas nos vinieron a la cabeza: ¿Es posible sintetizar bloques sólidos de alta pureza de este material? (El MgB_2 comercial es un polvo de poca pureza.) ¿Es realmente próxima a 40 K su transición superconductora? (Llevábamos casi dos decenios de OSNI, Objetos Superconductores No Identificados, compuestos de los que se había informado que presentaban altas temperaturas de transición, pero que nadie conseguía reproducir.) ¿Conseguiremos averiguar cuál es el mecanismo de la supuesta superconductividad? Y, por fin, ¿podremos esbozar algunas de las propiedades básicas del compuesto? Felizmente, la respuesta a todas estas cuestiones fue afirmativa.

El rumor del descubrimiento de Akimitsu desató un período de actividad frenética; resultó maravilloso para un gran número de equipos de investigación, entre ellos el nuestro, que está especializado en el estudio de las propiedades físicas de compuestos metálicos. Tan pronto como supimos del informe detuvimos todos los experimentos en curso y acometimos la producción de MgB_2.

La síntesis del diboruro de magnesio entrañó cierta dificultad inicial. Se trata de un compuesto intermetálico (combinación de dos o más elementos metálicos). El método más sencillo para obtener un producto de este tipo se basa en fundir conjuntamente los elementos que lo componen. Pero no fue posible aplicarlo en este caso ya que las temperaturas de fusión eran muy distintas: 650 ºC para el magnesio y mayor que 2000 ºC para el boro. Al hervir a unos pocos grados por encima de 1100 ºC, el magnesio se evaporaría antes que pudiera formarse el compuesto.

Sin embargo, la vaporización del magnesio sugirió un método alternativo. Introducimos un trozo de magnesio y un poco de polvo de boro en el interior de un recipiente de tántalo, que es inerte; tras sellarlo, lo sometimos a una temperatura suficientemente alta para que el magnesio se fundiera pero no se evaporara (unos 950 ºC). El magnesio tiene una presión de vapor relativamente elevada: a 950 ºC, un tercio de atmósfera de vapor de magnesio coexiste en equilibrio con el metal líquido. Esperábamos que este vapor denso se difundiera en el boro sólido, produciendo granos de MgB_2. En efecto, descubrimos que, tras sólo dos horas, este proceso producía MgB_2 de alta pureza en forma de granos o pellas ligeramente sinterizados (como arenisca). A los tres días de haber oído los rumores sobre el MgB_2, habíamos hallado un método para producirlo y confirmado su superconductividad a 40 K.

A continuación intentamos averiguar si se trataba de un superconductor clásico, es decir, si su actividad se explicaba mediante la teoría BCS (de las iniciales de los apellidos de sus tres autores: Bardeen, Cooper y Schrieffer), o si pertenecía a un tipo exótico. Si de lo último se tratara, las repercusiones serían las propias de un gran descubrimiento científico. Si resultaba ser un superconductor estándar tipo BCS, debíamos encontrar una nueva explicación a tan elevada temperatura de transición. Con todo, la perspectiva de desarrollar nuevas aplicaciones resultaba de lo más estimulante.

Por razones diversas, varios expertos pensaron que el MgB_2 no era un superconductor tipo BCS común. Antes de que se descubrieran los superconductores de alta temperatura crítica en 1986, la máxima temperatura de transición conocida había permanecido atascada durante dos decenios en alrededor de 20 K. Este hecho condujo a algunos teóricos a sugerir que 30 K era la máxima temperatura crítica posible en superconductores que obedecieran los principios de la teoría BCS. Los óxidos de cobre superconductores de alta temperatura crítica superaban con creces dicho límite, pero no se consideraban superconductores BCS.

Además, la temperatura de transición relativamente alta del MgB_2 violaba uno de los viejos principios que hasta entonces guiaban la búsqueda de compuestos intermetálicos con temperatura crítica (T_c) elevada: cuantos más electrones participen en la transición de fase al estado superconductor, mayor será T_c. Ni el magnesio ni el boro aportaban un gran número de electrones al MgB_2.

Una prueba experimental directa (basada en la sustitución isotópica) permite comprobar si un superconductor responde a la teoría BCS.

Resumen/El diboruro de magnesio

- En 2001 se descubrió que el MgB_2, considerado hasta entonces un compuesto anodino, se torna superconductor por debajo de unos 40 K: una temperatura crítica que casi dobla la de otros superconductores similares. Su temperatura operativa máxima se halla entre 20 y 30 K.
- Estas temperaturas pueden conseguirse enfriando el material con neón líquido, hidrógeno líquido o mediante un refrigerador de ciclo cerrado. Tales métodos criogénicos entrañan menores costes y menor complejidad técnica que la refrigeración con helio líquido que se utiliza para las aleaciones de niobio, de amplia aplicación en la industria a unos 4 K.
- Dopado con carbono u otras impurezas, el MgB_2 iguala, si no supera, las aleaciones de niobio en cuanto al mantenimiento de la superconductividad en presencia de un campo magnético y al transporte de corriente eléctrica. Las aplicaciones potenciales incluyen bobinas superconductoras (imanes), cables para el transporte de electricidad y detectores de campo magnético de gran sensibilidad.

Las vibraciones de la red atómica desempeñan una función clave en la teoría. Imaginemos que los iones pesados dotados de carga positiva de la red cristalina están sujetos por muelles (los enlaces químicos). Las excitaciones como las debidas al calor, por ejemplo, provocan la vibración de estos iones a frecuencias características. La teoría BCS predice que la temperatura de transición de un superconductor es proporcional a la frecuencia de las vibraciones de su red. Dado que los cuerpos de menor masa presentan frecuencias características superiores a las de otros cuerpos idénticos pero dotados de mayor masa (pensemos en objetos corrientes como unos vasos de vino o las cuerdas de una guitarra, por ejemplo), si usáramos un isótopo distinto de magnesio o boro obtendríamos MgB_2 con átomos de masa diferente que alterarían las vibraciones de red y, por tanto, modificarían la T_c.

El boro se presenta en la naturaleza en dos formas isotópicas estables: boro 10 y boro 11. Una versión simplificada del modelo BCS predice que la T_c debería diferir en 0,85 K en dos muestras de boro 10 o boro 11 puros. Nuestras primeras muestras sinterizadas de MgB_2 presentaban una diferencia de 1 K. Según la teoría BCS, esta ligera discrepancia entre el valor predicho y el resultado experimental indica que las vibraciones del boro son más importantes para la superconductividad que las del magnesio.

La proximidad de ambas cifras (teórica y experimental) sugiere que el MgB_2 corresponde a un superconductor BCS, si bien se trata de un caso extremo, con una temperatura de transición mayor que cualquier otro de este tipo. Parecía, pues, que la predicción de un límite superior de 30 K para la T_c de un superconductor BCS quedaba refutada. Ello era sin duda una buena noticia, ya que resulta más fácil trabajar con superconductores intermetálicos tipo BCS; por ejemplo, la fabricación de cables es más simple que con los superconductores basados en óxidos de cobre.

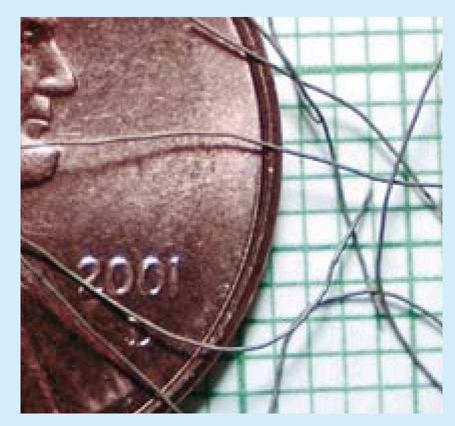
De hecho, caímos en la cuenta de que podían obtenerse cables de MgB_2 a través de la simple exposición de filamentos de boro a una atmósfera de vapor de magnesio. Dichos cables son de mayor utilidad que las pellas para el estudio de las propiedades del material y para aplicaciones tales como la construcción de imanes (bobinas generadoras de campo magnético).

Fabricación de cables

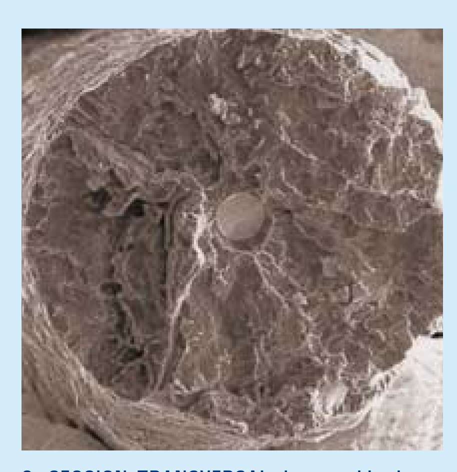
Sólo un par de semanas después del anuncio de la superconductividad del MgB_2 habíamos conseguido diseñar un método para producir segmentos de cable del nuevo superconductor. El MgB_2 se obtiene mediante la reacción de vapor de magnesio con boro: tras varias horas a temperaturas próximas a 1000 ºC, el boro absorbe el vapor de magnesio del entorno y se transforma en MgB_2. En este proceso, el boro se hincha de forma notable, lo mismo que una esponja seca que absorbiera vapor de agua del aire en un día húmedo. Para ello se compran fibras de boro de centenares de metros de longitud. Esta técnica se ha aplicado a filamentos con diámetros iniciales de entre 0,1 y 0,3 mm.

Tales segmentos son de gran utilidad para la determinación de las propiedades físicas del MgB_2. Para aplicaciones prácticas, sin embargo, necesitarían una funda conductora y maleable que prestase soporte estructural. (La funda conductora transportaría también la corriente eléctrica en caso de que el superconductor fallase, impidiendo así un calentamiento catastrófico del MgB_2.) Una funda así no se ha desarrollado todavía.

Un método más corriente para producir cables se denomina "polvo en un tubo". Consiste en rellenar un tubo con polvo de magnesio y boro o MgB_2 en polvo. Por esa vía se han obtenido muestras de una longitud comprendida entre decenas y centenares de metros.



1. CABLES producidos mediante la reacción de vapor de magnesio y filamentos de boro.



2. SECCION TRANSVERSAL de un cable de diboruro de magnesio roto. Revela en su interior un núcleo central de boruro de tungsteno de 0,015 mm de diámetro.

NORMAN E. ANDERSON, JR. (*arriba*); HAL SAILSBURY *Materials Preparation Center, Ames Laboratory* (*abajo*)

Aplicaciones de los superconductores

Aunque se manifiesta sólo a muy bajas temperaturas, la superconductividad ofrece una amplia variedad de aplicaciones, actuales y potenciales. Algunas de las más obvias derivan de la capacidad de los superconductores de soportar el paso de intensas corrientes eléctricas sin merma de energía ni calentamiento resistivo. Por ejemplo, imanes superconductores que producen campos magnéticos mayores que 20 tesla (alrededor de 500 veces más intenso que el producido por uno de los imanes que tenemos pegados a nuestra nevera). Imanes superconductores como éstos, así como otros de menor intensidad, se usan en laboratorios de investigación y en hospitales, en los equipos de obtención de imágenes por resonancia magnética nuclear. La verdad es que aumenta la venta de estos imanes, construidos a partir de compuestos y aleaciones de niobio.

También se ha propuesto la aplicación de los superconductores en cables de transporte de electricidad sin pérdida de energía, ya que pueden soportar mayores densidades de corriente que los no superconducto-

Historia de la superconductividad

Heike Kamerlingh Onnes descubrió la superconductividad en 1911 mientras utilizaba el helio líquido como refrigerante para estudiar las propiedades eléctricas de los metales a bajas temperaturas. Para sorpresa general, al enfriarse a unos 4,2 K, el mercurio perdía toda su resistencia eléctrica. Este valor umbral de transición al estado superconductor se denomina temperatura crítica, T_c.

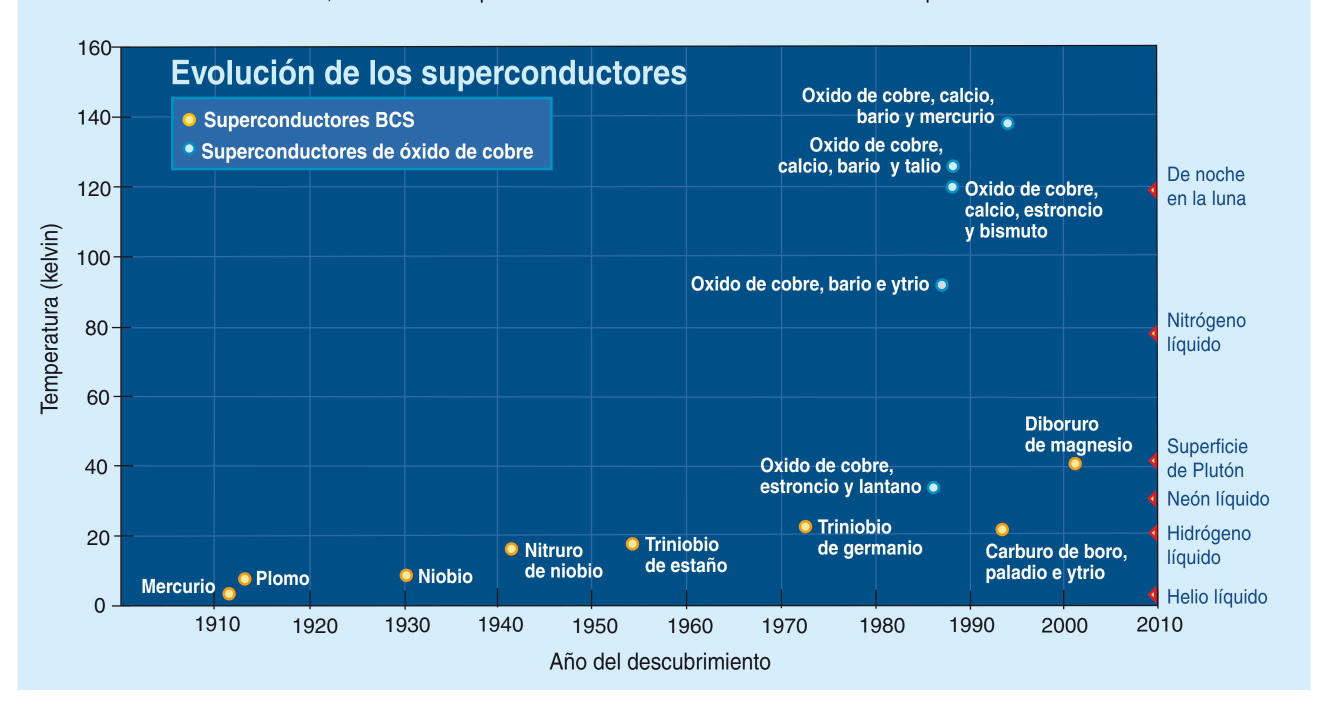
Otros superconductores de temperatura crítica incluso superior fueron descubriéndose, de forma lenta aunque firme, en el transcurso de los primeros cinco decenios de la investigación en superconductividad. Correspondían todos a elementos metálicos puros o bien a compuestos intermetálicos (combinaciones de dos o más elementos metálicos). Pero desde mediados de los sesenta hasta mediados de los ochenta el valor máximo de T_c parecía haberse estancado justo por encima de los 20 K.

La situación cambió en 1986 con el descubrimiento de la superconducción de alta temperatura en un gran número de compuestos de óxido de cobre. En los primeros años posteriores a este descubrimiento, los valores de T_c se dispararon, alcanzando el óxido de cobre, mercurio, bario y calcio una temperatura crítica de 130 K. Fue un período realmente excitante. Pero pronto se hizo patente que la teoría BCS, la más empleada para describir la superconductividad, no explicaba la ausencia de resistencia en estos materiales. A pesar de un esfuerzo de casi 20 años, todavía no disponemos de una teoría que explique por qué los compuestos de óxido de cobre son superconductores.

Estos compuestos han planteado múltiples retos. Al principio era difícil producirlos, en formas de alta pureza o monocristales; ello dificultaba la medición de sus propiedades fundamentales. Además, la obtención de cables no resulta fácil: a diferencia de los superconductores intermetálicos, los granos individuales que forman uno de estos óxidos deben estar alineados entre sí para que el cable ofrezca propiedades de interés en ingeniería. Debía encontrarse, pues, una sustancia con unas propiedades parecidas a las de los superconductores intermetálicos, sin dejar de poseer una temperatura crítica notablemente superior a 20 K.

Al amanecer del nuevo milenio, el estado superconductor podía alcanzarse de distintas formas, que variaban en coste y dificultad. La superconductividad en óxidos resultaba útil cerca de 77 K, temperatura que puede alcanzarse con bastante facilidad sumergiendo el material en nitrógeno líquido. Los compuestos intermetálicos más antiguos, tales como el triniobio de estaño, se utilizaban en el laboratorio y en imanes de uso en medicina que operaban a temperaturas próximas a 4 K (alcanzables mediante helio líquido).

En 2001 llegaría el material esperado: se descubrió que el diboruro de magnesio, un sencillo compuesto intermetálico, se tornaba superconductor a 40 K, temperatura crítica que casi dobla la de otros superconductores similares.



TOMMY MOORMAN Y CARLI MORGENSTEIN

res. Hasta la fecha, se han probado con éxito varios prototipos basados en óxidos de cobre, que han sido enfriados a unos 70 K con nitrógeno líquido.

En general, para que un superconductor opere como tal en aplicaciones prácticas es necesario enfriarlo varios grados por debajo de su T_c (aproximadamente entre 0,5 y 0,7 T_c) por una razón obvia: a una temperatura demasiado próxima a T_c, la superconductividad se destruye con corrientes eléctricas y campos magnéticos intensos. En consecuencia, una T_c de 20 K puede exigir una temperatura de trabajo de 10 K, lo que significa que el superconductor debe enfriarse con helio líquido, una operación costosa y algo compleja.

El MgB_2 atrae el interés de los que se dedican a la investigación aplicada porque se enfría a las temperaturas operativas con mayor facilidad que las aleaciones basadas en niobio (de T_c inferior) que se utilizan hoy. Puede reducirse la temperatura del MgB_2 con hidrógeno líquido, neón líquido o con un refrigerador de ciclo cerrado, bastante barato, que alcanza fácilmente temperaturas inferiores a 20 K.

Mas para que estos sueños se conviertan en realidad, es necesario que

el MgB_2 ofrezca mejores prestaciones superconductoras. Los expertos dedican especial atención a la fase superconductora mixta (en la que un campo magnético aplicado destruye la superconductividad parcialmente), ya que en la mayoría de las aplicaciones el material se encontrará en esta fase.

Con un campo magnético débil no se alcanza el estado mixto: el superconductor expele el campo magnético de su interior y se mantiene superconductor. Si el campo es intermedio, en cambio, consigue penetrar en el material en forma de pequeños tubos de flujo magnético denominados "vórtices". En el interior de éstos, el material pierde la superconductividad, pero en su exterior se mantiene superconductor.

Esta fase mixta mantiene gran parte de las propiedades superconductoras útiles. Al aumentar la intensidad del campo magnético aplicado, el porcentaje de material ocupado por los tubos de flujo aumenta hasta que éstos se superponen completamente, de forma que todo el material se torna no superconductor. La intensidad del campo a la que se pierde la superconductividad se denomina campo crítico superior; constituye una propiedad clave, pues determina la utilidad práctica del superconductor.

En su mayoría, las aplicaciones utilizarán campos magnéticos intermedios (suficientemente intensos para que el material resulte operativo, pero no hasta el punto de que destruyan la superconductividad). El objetivo consiste, pues, en maximizar los intervalos de temperatura y campo magnético en los que la fase superconductora mixta sobrevive. La temperatura desempeña también una función relevante en estas consideraciones, dado que el campo crítico de un superconductor depende de la misma. Justo por debajo de T_c, el campo crítico es casi cero, es decir, incluso el campo más débil destruye la superconductividad. A temperaturas inferiores, la superconductividad resiste campos de mayor intensidad.

Afortunadamente, el campo crítico superior de un material puede modificarse alterando la composición del mismo: en general, añadiendo impurezas. Por ejemplo, la sustitución de una parte del boro del MgB_2 por carbono mejora drásticamente su campo crítico superior. Nuestro grupo y otros han demostrado que con una sustitución del 5 por ciento de carbono se dobla el campo crítico superior del MgB_2, lo que supone una gran mejora en muestras masivas.

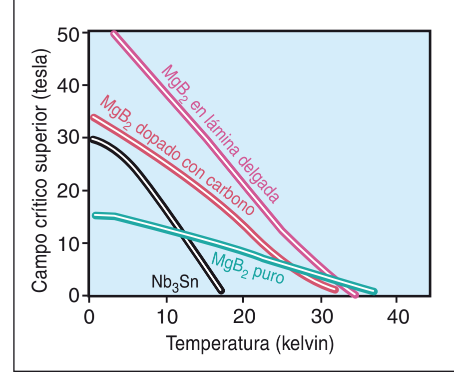
Además, el grupo de David C. Larbalestier, de la Universidad de Wisconsin-Madison, ha demostrado que las capas delgadas de MgB_2 presentan valores de campo crítico superior aún mayores, que superan incluso los del triniobio de estaño (Nb_3Sn). Estos datos sobre la capa delgada plantean una serie de cuestiones: ¿qué es lo que da lugar a tan altos valores? ¿Serán pequeñas cantidades de oxígeno? ¿Quizás algún otro elemento que se introduce y lo dopa de forma desconocida? ¿O se deben a las tensiones en la estructura del MgB_2 en la capa delgada? Cualquiera que sea la respuesta, el MgB_2 se presenta claramente como un material prometedor para la construcción de bobinas superconductoras (imanes) que operen a temperaturas superiores y quizás incluso bajo la acción de campos más intensos que el Nb_3Sn, el compuesto preferido en la actualidad para este tipo de imanes.

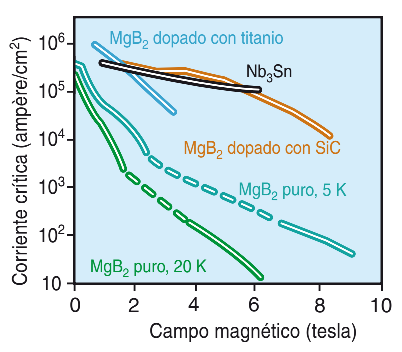
La segunda propiedad superconductora de especial interés para la física aplicada es la densidad de corriente crítica. Define ésta la máxima cantidad de corriente que puede transportar un superconductor manteniendo nula su resistividad. Cuando la densidad de corriente supera a la densidad de corriente crítica, los vórtices (las pequeñas regiones no superconductoras de la muestra) comienzan a moverse y se produce una perdida de energía: el material ofrece entonces una resistencia no nula. Para contrarrestar este efecto, los vórtices pueden ser anclados o clavados introduciendo el tipo adecuado de defecto (estructural) en el superconductor. A menudo, este anclaje puede potenciarse reduciendo el tamaño de los cristalitos individuales (o granos) del material, puesto que ello aumenta el área superficial asociada con las fronteras de grano, que es donde se anclan los vórtices. Otro método para aumentar el anclaje consiste en introducir inclusiones microscópicas de un segundo material, como el óxido de ytrio o el diboruro de titanio.

Incrementar la densidad de corriente crítica a campos magnéticos superiores constituye uno de los mayores

MEJORAR LAS PRESTACIONES

Mantener el estado superconductor en el seno de un campo magnético y al paso de una corriente eléctrica resulta crucial en las aplicaciones prácticas de los materiales superconductores. Los datos representados aquí muestran cómo el dopaje con impurezas mejora el comportamiento del MgB_2: consigue igualar o incluso exceder las prestaciones del triniobio de estaño (Nb_3Sn), el preferido por la industria. La gráfica de la izquierda muestra que un cable de MgB_2 dopado con carbono y una lámina delgada de MgB_2 con un nivel de impurezas desconocido soportan un campo magnético de mayor intensidad ("campo crítico superior") que el Nb_3Sn a cualquier temperatura. La gráfica de la derecha (tomada a 4 K, excepto donde se diga otra cosa) muestra que el MgB_2 dopado con carburo de silicio (SiC) iguala en capacidad de transporte de corriente eléctrica al Nb_3Sn, si bien otras variantes ofrecen peor comportamiento. (Las líneas discontinuas corresponden a interpolaciones.)



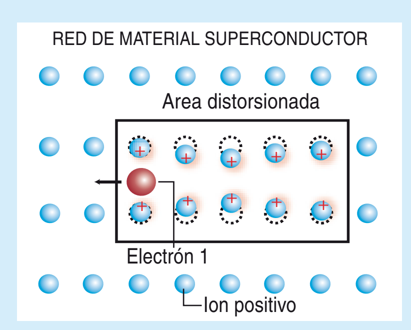


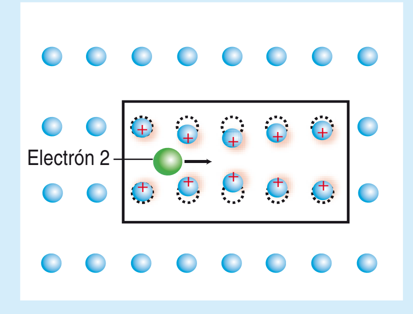
LUCY READING-IKKANDA

Predicciones de la teoría BCS

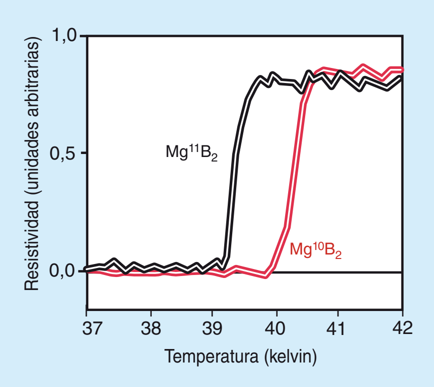
En 1957, los físicos John Bardeen, Leon N. Cooper y J. Robert Schrieffer propusieron un modelo para explicar el mecanismo que subyace en la superconductividad de los metales: la teoría BCS (denominada así por las iniciales de sus apellidos). En un metal no superconductor, los electrones son dispersados por defectos e imperfecciones estructurales; así se genera la resistencia. La superconductividad, en cambio, tiene lugar (de acuerdo con la teoría BCS) cuando los electrones se comportan como una sola entidad colectiva extensa que se mueve sin dispersarse.

Los elementos constructivos de este nuevo estado electrónico son los pares de Cooper: pares de electrones en los que cada uno es atraído ligeramente por el otro. Esta atracción entre dos partículas de carga idéntica —que a primera vista parece imposible— ocurre porque el metal está compuesto por iones dotados de carga positiva, además de por electrones. Cuando un miembro del par de Cooper se mueve a través del metal, deja en su estela una distorsión de carga positiva. Esta efímera carga neta positiva atrae al segundo electrón. De esta forma, la distorsión de la red acopla ligeramente a los electrones. (Para ser más precisos, en el acoplamiento están implicadas las vibraciones de red de una frecuencia determinada.) Podríamos compararlo al movimiento de dos niños en una cama elástica: aunque no existe atracción directa entre ellos, tienden a botar uno hacia el otro debido a la distorsión que producen en la lona bajo sus pies.





1. LA FORMACION DE PARES DE COOPER (pares de electrones acoplados) da lugar a la superconductividad. Al moverse un electrón a través de un metal, deja en su estela una distorsión de la red de iones dotados de carga positiva (*izquierda*). Poco después, el segundo electrón es atraído por la resultante concentración de carga positiva (*derecha*). De ello resulta una débil atracción mutua entre los dos electrones.



2. LA RESISTIVIDAD ELECTRICA del MgB_2 desaparece al enfriarse el material por debajo de su temperatura crítica, próxima a 40 K. La temperatura crítica de muestras sintetizadas con isótopos de boro puro 10 y 11 es distinta. Este desplazamiento isotópico queda explicado por la teoría BCS. De ello se infiere que la superconductividad del MgB_2 es la tradicional, del tipo BCS.

Los pares de Cooper se solapan entre sí. Por debajo de la temperatura crítica (T_c) forman un estado electrónico extenso que ya no encuentra resistencia eléctrica alguna.

Una versión simplificada de la teoría BCS predice que la T_c depende de tres propiedades del material: el número de electrones que participan en el estado superconductor (cuantos más electrones participen, mayor es T_c), la frecuencia característica de las vibraciones de red involucradas en el acoplamiento de los electrones del par de Cooper (a mayor frecuencia, mayor T_c) y la intensidad del acoplamiento entre la distorsión de red y los electrones (a mayor acoplamiento, mayor T_c). Durante decenios, la búsqueda de una T_c superior se centró en la optimización de estas tres propiedades relacionadas, sobre todo las dos primeras. El MgB_2 parece tener una T_c más alta en razón de un mayor acoplamiento electrón-red, es decir, en virtud de la tercera propiedad.

retos actuales para convertir el MgB_2 en un superconductor operativo. La densidad de corriente crítica del MgB_2 es comparable a la del Nb_3Sn a bajos campos magnéticos, pero decae más rápidamente a campos de mayor intensidad. Ello dificulta la utilización del MgB_2 en bobinas diseñadas para crear un campo magnético de gran intensidad. Con todo, desde el descubrimiento de la superconductividad en este compuesto hace cuatro años, se han logrado mejoras notables en la densidad de corriente crítica, en su valor a campos bajos y —quizá lo más importante— a campos altos. La investigación en este sector se encuentra en plena ebullición: parece que pronto se logrará optimizar las prestaciones superconductoras del MgB_2.

Pasado, presente y futuro

Resultado de decenios de investigación, el descubrimiento de la superconductividad en el MgB_2 nos recuerda que la naturaleza no siempre cumple los principios que establecemos en nuestros intentos —con frecuencia vanos— de describirla. Aunque el MgB_2 se descubrió hace más de 50 años, nunca se había comprobado si era superconductor, en parte porque no se ajustaba a nuestra idea de superconductor intermetálico. Afortunadamente, en la búsqueda de nuevos materiales y nuevas propiedades, la voz de la naturaleza todavía puede oírse a pesar del estruendo de nuestros prejuicios.

En los últimos cuatro años, la comprensión de la superconductividad del MgB_2 ha avanzado a velocidad de vértigo. Tenemos una clara idea de las propiedades del MgB_2 de alta pureza. Ahora estamos aprendiendo a modificar el material para mejorar los intervalos de campo magnético y densidad de corriente en que resulta útil. Las propiedades físicas a 20 o 30 K han mejorado hasta tal

LUCY READING-IKKANDA

punto, que podrían desarrollarse aplicaciones del tipo de densidad de corriente alta, así imanes (bobinas), que pueden refrigerarse con hidrógeno líquido, neón líquido o con un refrigerador de circuito cerrado. Ya se han conseguido prototipos de cables recubiertos e incluso algunas bobinas, pero todavía queda mucho trabajo por hacer: optimizar las propiedades de los superconductores y comprender mejor su metalurgia así como la de los posibles materiales de recubrimiento de los cables.

Con todo, el futuro del MgB_2 parece bastante prometedor. De producirse un cambio hacia una economía basada en el hidrógeno, el MgB_2 podría desarrollar su verdadero potencial. Si van a generarse grandes cantidades de hidrógeno, en los pequeños reactores de lecho granular [*véase* "Nueva generación de la energía nuclear", por James A. Lake, Ralph G. Bennett y John F. Kotek; INVESTIGACIÓN Y CIENCIA, marzo 2002], por ejemplo, ese hidrógeno debería transportarse de alguna manera. Para ello podría recurrirse a tubos aislados térmicamente que permitan transportar el líquido manteniendo la temperatura por debajo de la temperatura de ebullición del hidrógeno (20 K). Estas tuberías ofrecerían un sistema criogénico para cables de transporte de electricidad sin pérdidas hechos de MgB_2 que compartieran el espacio en el interior del tubo aislado térmicamente. (Aunque pueda sonar a fantasía científica, no a genuina ingeniería, tal sistema ya ha sido propuesto para su desarrollo.)

Tras el descubrimiento del primer superconductor de óxido de cobre, se han hallado decenas de otros óxidos de cobre superconductores. Sin embargo, cuatro años después del descubrimiento del MgB_2 no se ha encontrado ningún otro compuesto relacionado que muestre una T_c anormalmente elevada. El hallazgo de la superconductividad en los óxidos podría compararse con el descubrimiento de todo un continente provisto de inmensos espacios que explorar. La superconductividad del MgB_2, en cambio, vendría a ser como una isla algo separada de un archipiélago ya explorado. No sabemos si éste es el último eslabón de una cadena o si el futuro nos depara todavía más sorpresas.

LUCY READING-IKKANDA

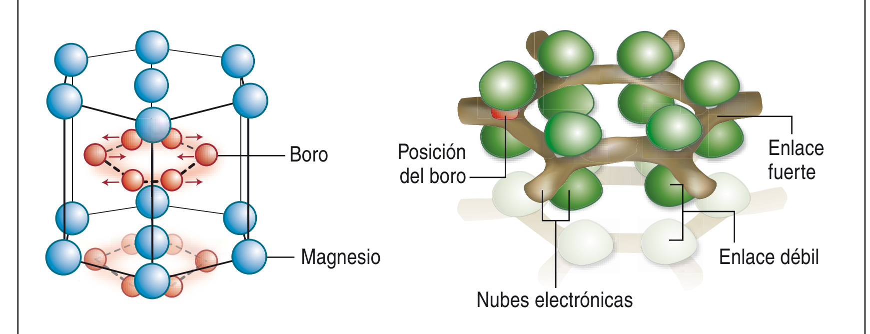
Estructura y enlaces

Una de las principales causas de la temperatura de transición sorprendentemente elevada del MgB_2 es la intensidad del acoplamiento entre los electrones y las vibraciones de red. Tal interacción surge de la especial estructura química del material.

En el MgB_2, los átomos de boro forman una red hexagonal en panel de abeja (*rojo, a la izquierda*). Estas capas de boro están separadas por capas de magnesio (*azul*). Los electrones responsables de la conductividad normal así como de la superconductividad están asociados a las capas de boro. Participan en uno de los dos tipos de enlace que hay en el material (*derecha*): los átomos de boro que integran la red hexagonal están unidos mediante un enlace fuerte; entre las distintas capas de boro existe un segundo enlace, más débil.

Los electrones de conducción de los enlaces planos son muy sensibles a las vibraciones en el plano (*flechas, izquierda*). Esta fuerte interacción, o acoplamiento, da lugar a un estado que permanece superconductor a temperaturas elevadas.

El MgB_2 ha revitalizado una pregunta básica de lo más excitante en física: ¿Es posible un tipo de superconductividad debida a dos colectividades de electrones (*verde y marrón*), que formen dos mares distintos de pares de Cooper? Los datos experimentales sugieren que el MgB_2 constituiría el primer ejemplo de este fenómeno.



Los autores

Paul C. Canfield y Sergey L. Bud'ko trabajan en el Laboratorio del Departamento de Energía en Ames, Iowa. Canfield es profesor de física y astronomía en la Universidad estatal de Iowa. Centra su investigación en el diseño, descubrimiento, crecimiento y caracterización de nuevos materiales y fenómenos, principalmente de compuestos metálicos a bajas temperaturas. Bud'ko estudia las propiedades termodinámicas, magnéticas y de transporte de nuevos materiales, las oscilaciones cuánticas en metales y semimetales, y las propiedades físicas de materiales en condiciones extremas que combinan altas presiones, campos magnéticos intensos y bajas temperaturas. Los autores reconocen agradecidamente su fructuosa colaboración con R. Wilke, D. Finnemore, C. Petrovic, G. Lapertot, M. Angst, R. Ribeiro y N. Anderson.

Bibliografía complementaria

SUPERCONDUCTIVITY AT 39 K IN MAGNESIUM DIBORIDE. Jun Nagamatsu et al. en *Nature*, vol. 410, págs. 63-64; 1 de marzo, 2001.

MAGNESIUM DIBORIDE: ONE YEAR ON. Paul C. Canfield y Sergey L. Bud'ko en *Physics World*, vol. 15, n.º 1, págs. 29-34; enero, 2002.

ENERGY FOR THE CITY OF THE FUTURE. Paul M. Grant en *Industrial Physicist*, vol. 8, n.º 1, págs. 22-25; febrero/marzo 2002.

MAGNESIUM DIBORIDE: BETTER LATE THAN NEVER. Paul C. Canfield y George W. Crabtree en *Physics Today*, vol. 56, n.º 3, págs. 34-40; marzo, 2003.

SUPERCONDUCTIVITY IN MgB_2: ELECTRONS, PHONONS AND VORTICES. Dirigido por Wai Kwok, George W. Crabtree, Sergey L. Bud'ko y Paul C. Canfield, en *Physica C*, vol. 385, n.os 1-2; marzo, 2003.

Calidad óptica del ojo

**Conocer la calidad óptica del ojo humano es importante en física, oftalmología, optometría, psicología...
Qué es, cómo se mide o cómo mejorarla se cuentan entre las cuestiones centrales que la ciencia empieza a desentrañar**

Susana Marcos

El ojo es un instrumento óptico que proyecta las imágenes del mundo exterior sobre la retina. Aunque suele comparársele con una cámara fotográfica, la calidad óptica de sus lentes, la córnea y cristalino, es, en general, peor que la de un objetivo fotográfico común.

Si la anatomía y fisiología del ojo revisten máximo interés, no es menor la atracción que despierta para el físico. Se pregunta éste cómo evaluamos la calidad de las imágenes retinianas, cuáles son los factores que contribuyen a la degradación de la calidad óptica del ojo, qué condiciones se asocian a una mayor degradación de la calidad óptica ocular, cómo se ve alterada por cirugía corneal o intraocular, cómo se coordinan los componentes ópticos oculares para mejorar la calidad de imagen o la posibilidad de corrección de las imperfecciones oculares. De esa gavilla de cuestiones nos ocupamos aquí.

Las imperfecciones oculares predominantes son el desenfoque, que caracteriza la miopía, hipermetropía y el astigmatismo. Tales patologías se corrigen con gafas o lentillas. Pero el ojo sufre, además, otras imperfecciones ópticas, llamadas aberraciones de alto orden, que no se miden en la práctica clínica, ni pueden corregirse mediante los métodos al uso.

Esas aberraciones, igual que el desenfoque, producen un emborronamiento de la imagen retiniana: reducen el contraste y limitan el intervalo de frecuencias espaciales de la imagen, es decir, el grado de detalle. Reparemos en la figura 1. Se muestra allí un ejemplo de la proyección en la retina de la imagen de una fuente puntual, un patrón de franjas verticales y una escena de flamencos, observada a unos 100 m de distancia, para un ojo corregido de desenfoque y astigmatismo. La imagen de un punto se ensancha y se torna asimétrica por efecto de las aberraciones; disminuye el contraste de las franjas, tanto más cuanto mayor es la frecuencia espacial; en general, las imágenes pierden nitidez.

La contribución de las aberraciones de alto orden a la degradación óptica es, en general, menor que la del desenfoque o el astigmatismo. Su efecto se hace más evidente para pupilas de mayor diámetro; por ejemplo, por la noche. Para pupilas pequeñas, los efectos de difracción, asociados al tamaño limitado de la apertura, predominan sobre las aberraciones (*figura 2*).

Aparte de la difracción y de las aberraciones, la difusión (“scattering”) contribuye también a la degradación de la imagen retiniana. La difusión se produce principalmente en el cristalino. Es prácticamente despreciable en los

jóvenes, pero aumenta con la edad; provoca una pérdida notable de transparencia del cristalino en pacientes con cataratas.

Calidad de imagen retiniana

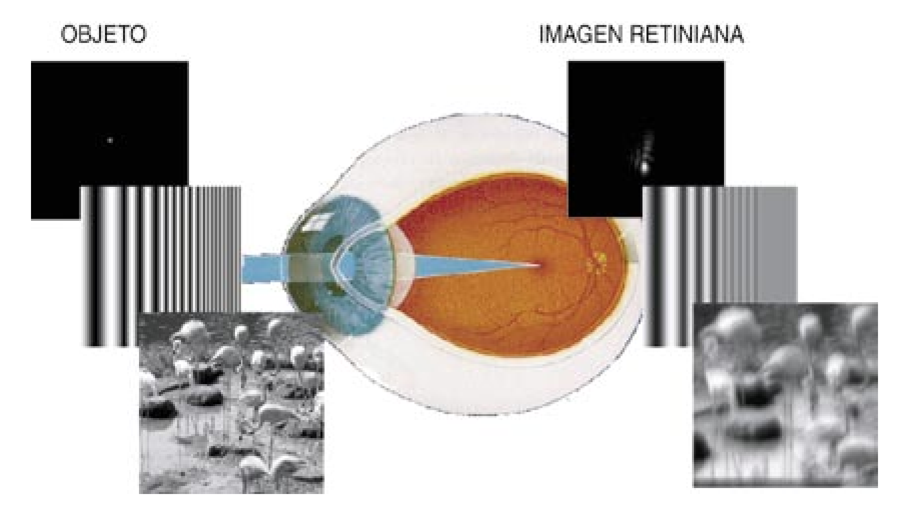
En ingeniería óptica es frecuente evaluar la calidad de imagen de un sistema óptico mediante el cálculo de las desviaciones que sufren rayos paralelos con respecto a las trayectorias ideales. En un sistema óptico perfecto (limitado por difracción), todos los rayos que entran por la pupila inciden en el mismo punto en el plano imagen, que, en el caso del ojo, se trataría de la retina.

Las desviaciones angulares con respecto a la trayectoria ideal se denominan aberraciones transversales. La representación habitual de las aberraciones de un sistema óptico se realiza en términos de la aberración de frente de onda; de ahí el tecnicismo “aberración de onda”. El frente de onda es normal a las trayectorias de los rayos. Conocidas las aberraciones transversales para cada posición de entrada del rayo en la pupila, que son proporcionales a las derivadas locales de la aberración de onda, es posible estimar la aberración de onda.

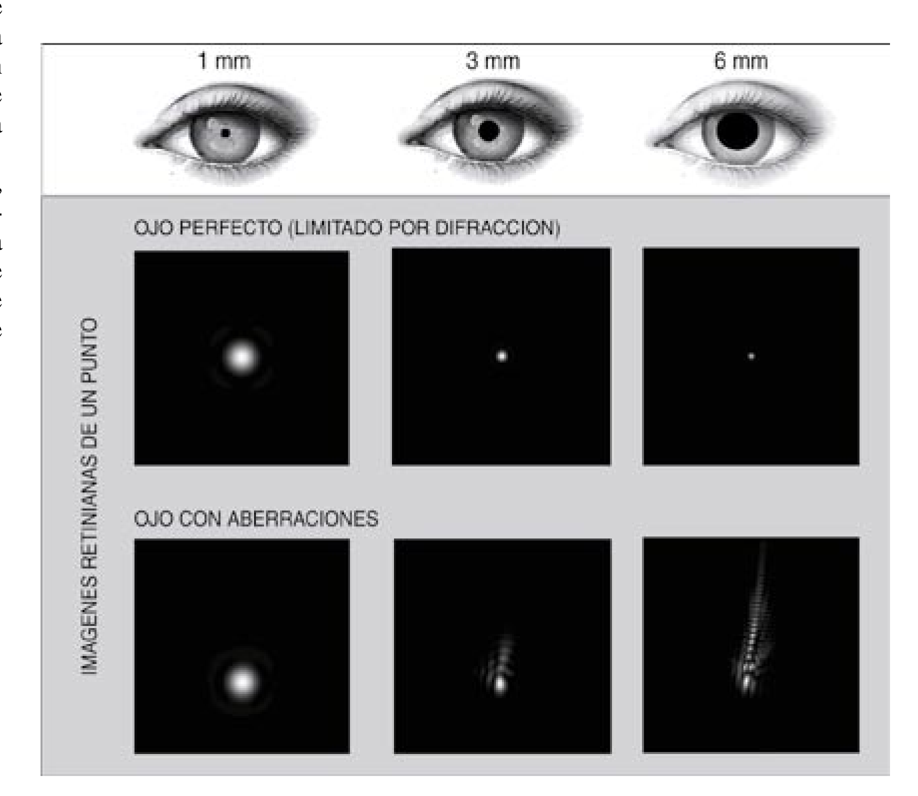
Un sistema óptico perfecto transforma un frente de onda plano en un frente de onda esférico, que converge en la imagen de un punto limitada por difracción. Un sistema óptico con aberraciones distorsiona el frente de onda con respecto al frente de onda esférico.

Para un ojo perfecto —esto es, sin aberraciones o limitado por difracción—, la aberración de onda es plana. Por aberración de onda se entiende la diferencia entre el frente de onda distorsionado y el frente de onda ideal. El error cuadrático medio del frente de onda (RMS) constituye, así, una métrica global de calidad óptica del ojo. La aberración de onda es una superficie compleja, que se suele describir como suma de polinomios más simples (expansión en polinomios de Zernike).

Los coeficientes de Zernike de orden inferior remiten a errores de refracción. Así, los coeficientes de primer orden representan un prisma; los de segundo, un desenfoque, que caracteriza a la miopía, hipermetropía y astigmatismo. Los términos de alto orden incluyen la aberración esférica, que indica un desenfoque variable con el diámetro de la pupila y produce halos en las imágenes, y el coma, que

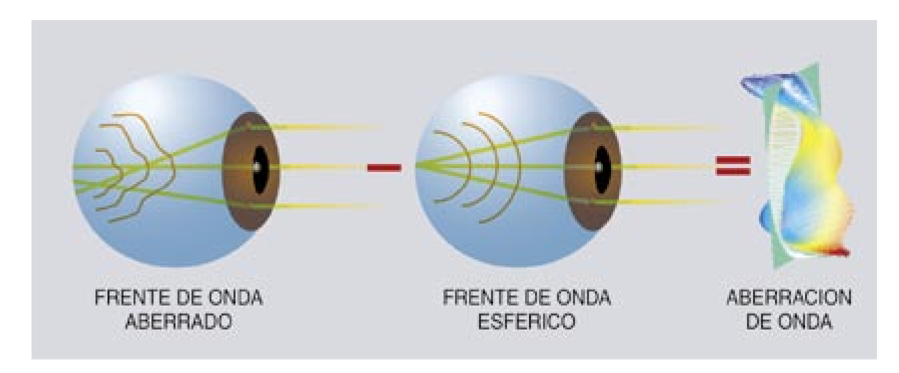


1. EL OJO PROYECTA LAS IMAGENES del mundo exterior sobre la retina. La figura muestra el efecto de las aberraciones oculares de un sujeto sobre la imagen de una fuente puntual, sobre un patrón de franjas verticales y sobre una escena natural. La óptica del ojo produce un emborronamiento y distorsión de las imágenes; disminuye el contraste, tanto más cuanto más alta es la frecuencia espacial (inverso del intervalo espacial entre franjas).



2. IMAGENES DE UNA FUENTE PUNTUAL (PSF) en la retina, para un ojo perfecto (*fila superior*) y para un ojo con aberraciones (*fila inferior*), para distintos diámetros de pupila. En un sistema sin aberraciones, la calidad óptica mejora al aumentar el diámetro de pupila: la imagen de un punto es más próxima a un punto. Un ojo real, con aberraciones, está limitado por difracción para pupilas de 1 mm, pero el impacto de las aberraciones crece al aumentar el diámetro de pupila.

SUSANA MARCOS



3. CONCEPTO DE ABERRACION DE ONDA. En un ojo limitado por difracción, un frente de onda plano converge hacia la retina como un frente de onda esférico. Las trayectorias de los rayos que entran en el ojo son perpendiculares al frente de onda. Todos los rayos, independientemente del lugar por donde entren en la pupila, convergen en la misma posición en la retina. En un ojo con aberraciones, el frente de onda que viaja hacia la retina se encuentra distorsionado y los rayos inciden en distintas zonas de la retina. Se llama aberración de onda a la diferencia entre el frente de onda distorsionado y el frente de onda esférico del sistema perfecto. Se puede obtener a partir de las desviaciones de los rayos con respecto a la trayectoria ideal.

es una aberración de tercer orden, sin simetría rotacional y produce imágenes dobles.

A partir de la aberración de onda podemos calcular la imagen de un punto en la retina (PSF) y la función de transferencia de modulación (MTF). La PSF designa la distribución de intensidades de la imagen de una fuente tras su paso por un sistema óptico. La MTF nos ofrece el grado de detalle, esto es, la reducción del contraste en función de la frecuencia espacial de la imagen a su paso por un sistema óptico.

El ojo humano es un filtro de paso bajo. Ello significa que la reducción del contraste es mayor para frecuencias espaciales altas (detalles finos en la imagen). Para describir la calidad de un sistema óptico suele recurrirse a la razón de strehl, una métrica que representa el máximo de la imagen de un punto en la retina del sistema óptico con relación al máximo de la PSF de un sistema sin aberraciones; equivalentemente, representa el volumen de la MTF con relación al volumen de la MTF del sistema sin aberraciones. Se considera que el sistema está limitado por difracción si la razón de strehl es superior a 0,8 (criterio de Raleigh).

ABERRACION DE ONDA	Distorsiones del frente de onda en el plano de la pupila. Las derivadas locales de la aberración de onda son proporcionales a las desviaciones de los rayos	$W(x,y)$	
FUNCION PUPILA	Función cuyo módulo es la transmitancia de la pupila, y su fase la aberración de onda	$P(x,y)=M(x,y)\cdot e^{-i\frac{2\pi}{\lambda}W(x,y)}$	
IMAGEN DE UN PUNTO (PSF)	Distribución de intensidades de la imagen de una fuente puntual tras su paso por un sistema óptico	$PSF=\lvert FT(P(x,y))\rvert^2$	
FUNCION DE TRANSFERENCIA DE MODULACION (MTF)	Reducción del contraste, en función de la frecuencia espacial de las imágenes a su paso por un sistema óptico	$MTF=\lvert IFT(PSF)\rvert$	
FUNCION STILES-CRAWFORD	Transmitancia gaussiana de la pupila, consecuencia de las propiedades direccionales (como guías de onda) de los fotorreceptores retinianos	$P(x,y)=M(x,y)\cdot e^{-i\frac{2\pi}{\lambda}W(x,y)}$	

4. DEFINICIONES, EXPRESIONES MATEMATICAS E ILUSTRACIONES de las funciones que se utilizan comúnmente para describir la calidad óptica del ojo.

Pupila y calidad óptica

La pupila del ojo regula la cantidad de luz que llega a la retina, igual que el iris en una cámara fotográfica. Además, el diámetro de la pupila determina la contribución de difracción y aberraciones a la calidad de imagen retiniana. Para pupilas pequeñas, la calidad óptica depende de la difracción; aunque el efecto de las aberraciones es pequeño, la resolución espacial es baja. Para pupilas grandes, la contribución de las aberraciones domina sobre la difracción.

El tamaño óptimo de pupila que proporciona la mejor calidad óptica depende de la magnitud y distribución de las aberraciones en cada ojo; ese tamaño óptimo se sitúa, en promedio, en torno a 3 milímetros. La mejor calidad de imagen se obtendría con un ojo ideal libre de aberraciones y el mayor diámetro de pupila; es decir, en torno a 8 milímetros, propio de un individuo joven.

El diámetro de pupila también determina, en parte, la profundidad de foco del ojo. A diferencia de lo que ocurre en un sistema limitado por difracción, en el cual la profundidad de foco decrece sistemáticamente al incrementar el diámetro de la pupila, la presencia de aberraciones tiende a igualar la profundidad de foco, que tiene un valor en torno a 0,3 dioptrías a partir de 4 milímetros de pupila.

Debido a las propiedades direccionales de los fotorreceptores retinianos, la pupila no transmite la luz homogéneamente en toda su área. Esas células sensibles a la luz se hallan orientadas hacia una misma posición en la pupila y se comportan como fibras ópticas; absorben de manera más eficiente la luz que incide a lo largo de sus ejes; es decir, la luz que entra en el ojo por la posición hacia la que se orientan. Este efecto, cuya expresión analítica aparece en la figura 4, abajo, se conoce como efecto Stiles-Crawford, y resulta en una pupila efectiva con transmitancia gaussiana. Adviértase que el pico del efecto Stiles-Crawford no es necesa-

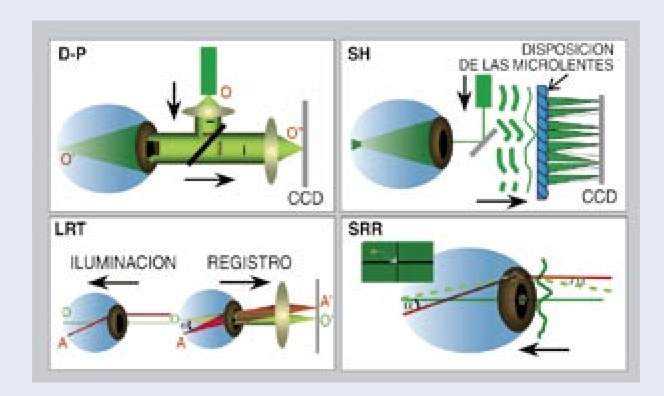
SUSANA MARCOS

METODOS PARA MEDIR LAS ABERRACIONES DEL SISTEMA OPTICO DEL OJO

CONCEPTOS BASICOS que subyacen bajo las técnicas para medir la calidad óptica y aberraciones del ojo. En la llamada técnica de doble paso (D-P), se proyecta una fuente puntual en la retina. La imagen de vuelta se forma, tras pasar dos veces por los medios oculares, sobre una cámara CCD. La función de transferencia de modulación del ojo se puede obtener directamente de dicha imagen aérea.

En la técnica de Shack-Hartmann (SH) se proyecta una fuente puntual sobre la retina; el frente de onda se muestrea, a la salida, con una matriz de microlentes. Un frente de onda perfecto forma múltiples imágenes aéreas en el punto focal de las microlentes. Para un frente de onda distorsionado, cada microlente muestrea una porción inclinada del frente de onda, de modo que las correspondientes imágenes se desvían de las posiciones ideales. A partir de estas desviaciones, se estima la aberración de onda en función de la posición en la pupila.

Las técnicas de trazado de rayos por láser ("Laser Ray Tracing", o LRT) y refractómetro de resolución espacial ("Spatially Resolved Refractometer", o SRR) son secuenciales. Muestrean el frente de onda en el camino de entrada hacia el ojo. En ambos aberrómetros, los rayos de luz entran en el ojo por distintas partes de la pupila. Por efecto de las aberraciones, los rayos se desvían con respecto al correspondiente a una pupila centrada. En el LRT, se captan imágenes sobre una cámara CCD, a medida que un haz láser escanea el ojo, mientras que en el SRR es el propio sujeto el que alinea el haz con respecto de una referencia usando una palanca. La aberración de onda se estima a partir de estas desviaciones en función de la posición en la pupila.



riamente el centro de la pupila; en muchos sujetos, los conos se orientan hacia una posición excéntrica en la pupila. En general, el efecto Stiles-Crawford, significativo sólo para pupilas grandes, amortigua el impacto de las aberraciones oculares.

Medición de la calidad de imagen

La presencia de aberraciones en el ojo se reconoció hace ya cerca de dos siglos. Merced a los avances técnicos, podemos ahora medir las aberraciones con rapidez y precisión. Los sistemas desarrollados y utilizados experimentalmente en laboratorios de investigación comienzan a abrirse ahora paso en la práctica clínica (*véase el recuadro*).

La función de transferencia de modulación del ojo se ha venido midiendo desde hace más de dos décadas, mediante la llamada técnica de doble paso. Consiste en la proyección de una fuente puntual láser en la retina y la captación de la imagen reflejada por la retina en una cámara CCD, colocada en un plano conjugado. La imagen aérea captada por el CCD contiene doble degradación por los medios oculares, en el camino de entrada y en el de salida del ojo; en términos matemáticos es la autocorrelación de la imagen de un punto. La MTF se puede estimar fácilmente a partir de la imagen aérea. Esta función de transferencia de modulación incorpora la degradación impuesta por las aberraciones monocromáticas, la difracción y la difusión.

La mayoría de los aberrómetros cuantifican la aberración transversal en función de la posición en la pupila. Podemos medir la aberración transversal durante el camino de entrada al ojo; para ello se emplea el llamado sistema de trazado de rayos por láser (LRT) o el refractómetro de resolución espacial (SRR). En el LRT la pupila se muestrea secuencialmente, mediante un sistema escáner que barre la pupila y proyecta la imagen de un punto en la retina. Una cámara capta las imágenes aéreas para distintas posiciones de entrada del haz en la pupila. Por efecto de las aberraciones, las imágenes que se proyectan a través de proyecciones excéntricas de la pupila se desvían con respecto del rayo central.

El SRR constituye otra técnica secuencial. En lugar de la captación de imágenes aéreas en una cámara CCD, el sujeto alinea las imágenes de un punto, observadas a través de las distintas posiciones en la pupila, con un estímulo de referencia, proyectado a través de una pupila centrada.

En la técnica de Shack-Hartmann (SH) se proyecta una fuente puntual sobre la retina. El frente de onda reflejado atraviesa una matriz de microlentes que focaliza el haz en múltiples imágenes del punto (una por cada lente) en una cámara. Cada lentecilla muestrea una pequeña porción del haz, correspondiente a una región de la pupila. Las desviaciones de los centroides de las imágenes aéreas del punto con respecto al punto focal de cada lentecilla representan la aberración local transversal de rayo. A partir de tales desviaciones se estima la aberración de onda.

Los cuatro métodos precedentes miden las aberraciones del todo el sistema óptico del ojo. Haciendo uso de un sistema de topografía corneal, podemos calcular las aberraciones de la córnea. La aberración transversal de la cara anterior de la córnea se obtiene mediante un trazado de rayos virtual sobre el mapa de elevación de la cornea; con tales datos se calcula la aberración de onda corneal.

Las aberraciones de los medios oculares internos, el cristalino por ejemplo, se obtiene sustrayendo la componente corneal de la aberración total.

Calidad de imagen retiniana policromática

Las técnicas que acabamos de describir utilizan luz monocromática. Miden, por tanto, la calidad de imagen monocromática. Sin embargo, el en-

SUSANA MARCOS



5. SISTEMAS EXPERIMENTALES con los que ha trabajado la autora. Tres se hallan en el Instituto de Optica (CSIC) de Madrid: doble-paso (D-P), Shack-Hartmann (SH) y trazado de rayos por láser (LRT). El cuarto, un refractómetro de resolución espacial (SRR), se encuentra en el Instituto Schepens de Investigación Ocular, de la Universidad de Harvard.

torno visual es policromático, lo que significa que las aberraciones cromáticas, consecuencia de la dependencia del índice de refracción de los medios oculares con la longitud de onda, desempeñan un papel importante en la calidad de imagen.

Hay dos aberraciones cromáticas principales: longitudinal y transversal. La aberración cromática longitudinal (LCA) representa el cambio de foco con la longitud de onda; la aberración cromática transversal (TCA) se refiere a la dependencia de la desviación angular con la longitud de onda.

Los numerosos estudios que han medido la LCA mediante técnicas subjetivas o mediante técnicas de doble paso revelan que apenas existen diferencias notables entre individuos; a tenor de los mismos, la diferencia cromática de foco promedio entre el azul (longitud de onda de 450 nanómetros) y el rojo (650 nanómetros) es de 1,4 dioptrías.

Para la medida de la aberración cromática transversal foveal suele recurrirse a la aplicación de técnicas subjetivas de alineamiento de estímulos. Típicamente la TCA es distinta de cero en el centro de la pupila. En nuestro laboratorio hemos empleado una extensión de la técnica SRR en varias longitudes de onda, para medir la LCA (término de desenfoque en función de la longitud de onda), la TCA (desviación angular entre los puntos rojos y azules) y la variación de las aberraciones con la longitud de onda.

Con los datos obtenidos, se generan la imagen de un punto y la función de transferencia de modulación policromática. La figura 7 muestra la MTF monocromática y policromática para un sujeto. Se ha demostrado que, para un sistema óptico perfecto, sin aberraciones monocromáticas, el impacto relativo de las aberraciones cromáticas sobre la calidad óptica es mayor que para un sistema con aberraciones; se da, además, cierto grado de equilibrio entre las aberraciones cromáticas y monocromáticas del ojo.

Calidad de imagen retiniana en el ojo normal

Encontramos una amplia variación, en magnitud y en distribución, de las aberraciones oculares entre la población. En la figura 7 se nos ofrecen ejemplos de mapas de aberración de onda (previa cancelación del término de desenfoque) en un grupo de ojos normales. Las aberraciones tienden a presentar simetría especular entre los ojos izquierdo y derecho, si bien no son infrecuentes los sujetos con patrones de aberraciones bilaterales muy distintos.

La córnea y el cristalino son los principales componentes refractivos del ojo. Las aberraciones de cada uno de ellos contribuyen a la calidad de imagen retiniana. Se ha demostrado que, al menos en sujetos jóvenes, una parte de las aberraciones corneales está compensada por las aberraciones del cristalino. Además, la aberración esférica de la córnea tiende a ser positiva y la aberración esférica del cristalino a

ABERRACION CORNEAL
+
=
ABERRACION TOTAL
ABERRACION INTERNA

6. CALCULO DE LAS ABERRACIONES OPTICAS de los componentes intraoculares. Constituyen éstas la diferencia de las aberraciones totales menos las aberraciones corneales.

SUSANA MARCOS

ser negativa, al menos en sujetos amétropes y miopes.

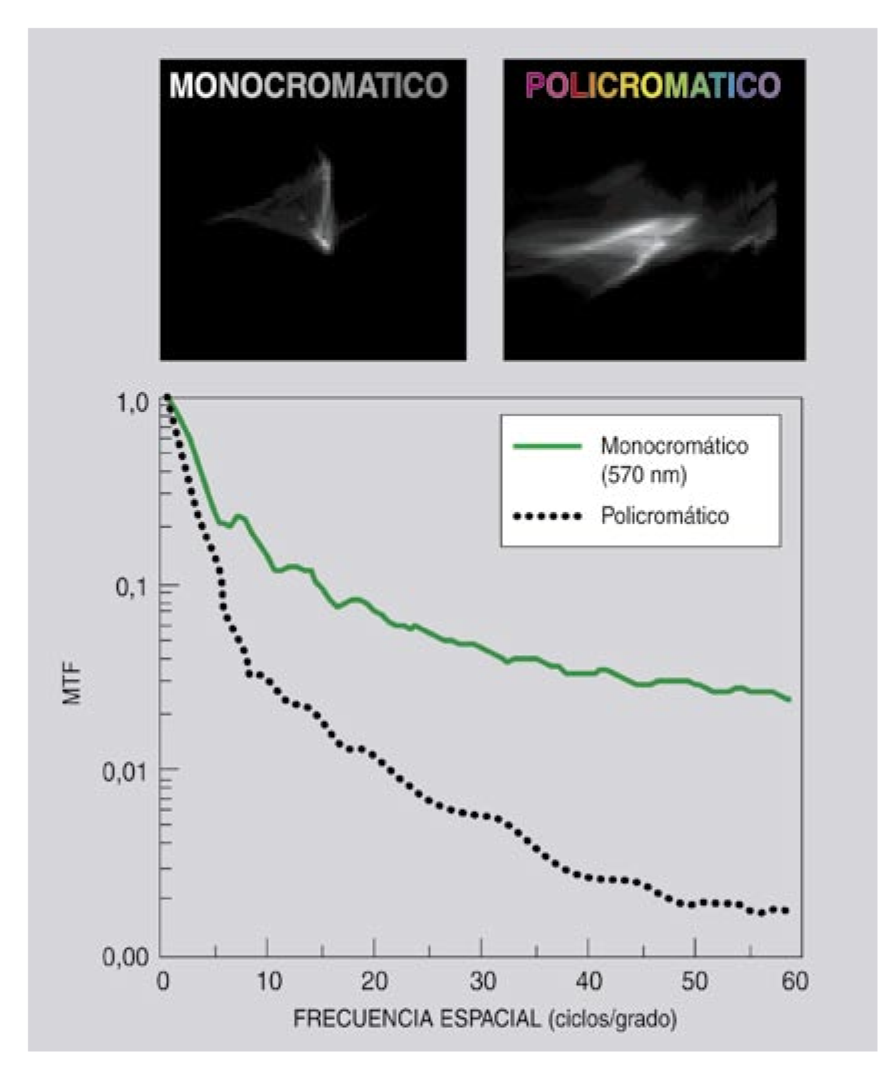
A pesar de que las medidas de aberraciones evaluadas en las mismas condiciones en un individuo son muy reproducibles, las aberraciones no son estáticas. Las aberraciones se modifican con el envejecimiento. La figura 9 muestra los cambios de las aberraciones oculares con el proceso de acomodación (enfoque de lejos y cerca), y con el envejecimiento. Las modificaciones de las aberraciones con el proceso de acomodación resultan de los cambios experimentados en la curvatura y posición del cristalino. El error cuadrático medio del frente de onda (RMS) alcanza un mínimo cerca del estado de reposo (en torno a 6 diámetros de esfuerzo acomodativo) y se incrementa para estímulos más cercanos. La aberración esférica varía sistemáticamente con la acomodación, igual que los términos de quinto orden y superior.

El envejecimiento produce una degradación de la óptica ocular. Esa disminución, no sólo deriva del incremento de la difusión en el cristalino, que puede desembocar en la formación de cataratas, sino que guarda relación también, con un incremento de las aberraciones. De la investigación realizada entre grupos de sujetos de distintas edades se desprende la existencia de un incremento de las aberraciones de alto orden con la edad, en particular de la aberración esférica y las de quinto orden y superior. Las mediciones del cristalino realizadas *in vitro* muestran un incremento de la aberración esférica hacia valores más positivos, aumento que altera la compensación de las aberraciones corneales e internas con la edad.

Algunas condiciones oculares llevan asociadas una mayor presencia de aberraciones. Se han observado, a este respecto, correlaciones significativas entre la miopía y la RMS de las aberraciones de alto orden (*véase figura 9*). El queratocono, una patología que se caracteriza por una protrusión anterior de la córnea, produce un incremento notable de la aberración de coma.

Pero los trabajos a que hemos aludido se refieren a medidas realizadas con fijación foveal. El ojo es un sistema imperfecto en eje; sin embargo, a diferencia de la mayoría de los instrumentos ópticos, la calidad no se degrada drásticamente fuera de eje. Las mediciones de aberraciones para distintas excentricidades evidencian un incremento sistemático del astigmatismo y el coma con el campo visual. Con todo, el decremento de la calidad de imagen no sigue la rápida disminución de la resolución del muestreo retiniano de la imagen.

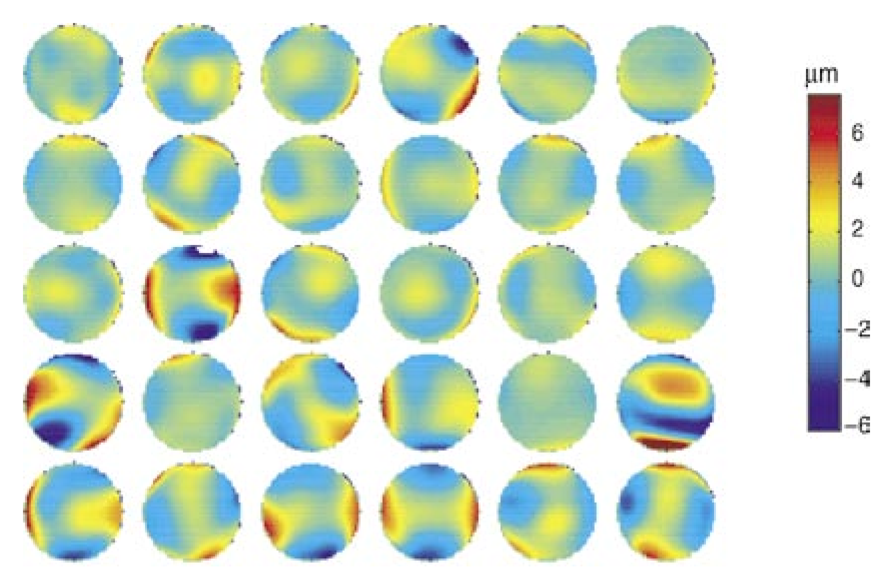
SUSANA MARCOS



7. IMAGENES PUNTUALES, MONOCROMATICA Y POLICROMATICA, para un sujeto normal, de mediana edad, y sus correspondientes funciones de transferencia de modulación (perfiles radiales). Las aberraciones de onda se midieron para distintas longitudes de onda (entre 450 y 650 nanómetros) utilizando la técnica del refractómetro de resolución espacial. La aberración cromática longitudinal (LCA) y la aberración cromática transversal (TCA) se obtuvieron mediante el mismo conjunto de datos. Las imágenes de un punto para cada longitud de onda se extrajeron a partir de cada aberración de onda, para 6 mm de pupila. La imagen de un punto policromática se logró como superposición de las imágenes de un punto monocromática, pesadas por la sensibilidad espectral de la retina. Se consideró el mejor foco para una longitud de onda de 570 nm; para el resto de las longitudes de onda se impuso el desenfoque producido por la LCA. Cada imagen de un punto monocromática se desplazó según la TCA.

Cirugía corneal y de cataratas

La cirugía corneal modifica la curvatura de la parte central de la córnea para corregir errores refractivos. La cirugía de cataratas sustituye el cristalino del ojo por una lente intraocular transparente. La modificación de la córnea y el cristalino conlleva una modificación de la calidad de imagen retiniana.



8. EJEMPLOS DE ABERRACION de onda para un grupo de sujetos normales.

La cirugía refractiva corneal se ha convertido en una alternativa frecuente a la corrección de los errores refractivos. Por estudios basados en la topografía corneal sabemos que, a pesar de que el desenfoque o el astigmatismo suelen corregirse con éxito, la cirugía refractiva provoca un incremento de las aberraciones corneales de alto orden.

Las mediciones de las aberraciones totales en pacientes, antes y después de cirugía refractiva LASIK, revelan un incremento de un factor 2 para el error cuadrático medio del frente de onda de aberraciones de tercer orden y superior por un factor 2 y de un factor 4 para la aberración esférica. La figura 10 (*panel superior*) muestra las aberraciones promedio de tercer orden y superior en un grupo de 14 sujetos antes y después de cirugía refractiva LASIK y un ejemplo de patrones de aberraciones totales y corneales de alto orden para un sujeto.

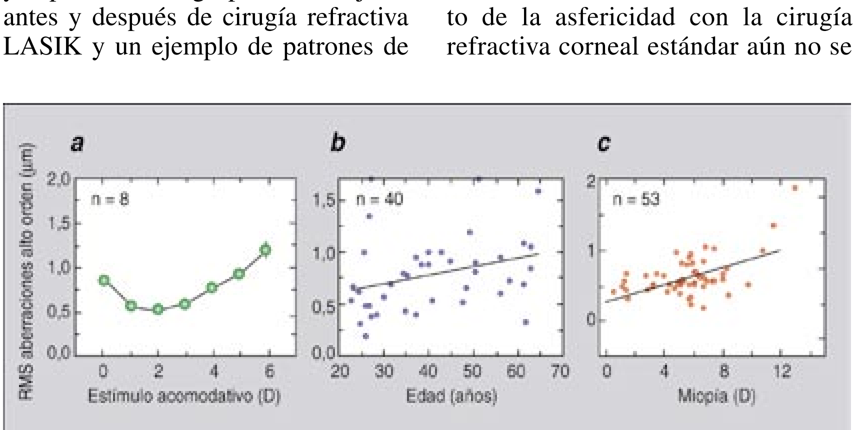
El mayor incremento en la RMS se produce sobre las aberraciones de la cara anterior de la córnea. Los cambios operados en las aberraciones internas indican una ligera modificación de la cara posterior corneal con la cirugía. La degradación del contraste producida por el incremento de aberraciones, medida en términos de la frecuencia espacial de la imagen, guarda correlación con una degradación de la sensibilidad al contraste medida subjetivamente en estos sujetos. El incremento de la aberración esférica guarda, a su vez, una estrecha correlación con el error esférico corregido: se asocia a un incremento de la asfericidad corneal. Pese a todo, las causas del incremento de la asfericidad con la cirugía refractiva corneal estándar aún no se conocen completamente. Para evitar la inducción de aberraciones de alto orden, parece preciso un refinamiento de los algoritmos de ablación y una mayor comprensión de la interacción láser-tejido corneal y de la respuesta biomecánica de la córnea.

En la cirugía de catarata se reemplaza el cristalino natural por una lente intraocular. Los fabricantes de lentes intraoculares evalúan la calidad de dichas lentes a través de medidas de su MTF en un banco óptico. La primera evaluación *in vivo* de la calidad de imagen retiniana en pacientes operados de cataratas se llevó a cabo mediante la técnica de doble-paso. Tras la cirugía, la frecuencia espacial de la imagen es significativamente más baja que la normal en sujetos jóvenes; y peor en pacientes con lentes multifocales que con lentes monofocales. En nuestro laboratorio del Instituto de Optica hemos medido por primera vez las aberraciones de la lente intraocular *in vivo*, sustrayendo las aberraciones corneales de las aberraciones totales en pacientes con lentes intraoculares implantadas.

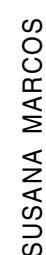
Exponemos en la figura 10 (*panel inferior*) el promedio de aberraciones totales, corneales y del cristalino en un grupo de pacientes y sujetos control; se ofrece también un caso de aberración de onda de un paciente operado en comparación con un sujeto joven. Las aberraciones corneales tienden a aumentar, probablemente por efecto de la incisión. Además, las mediciones *in vitro* revelan que las lentes intraoculares esféricas exhiben aberración esférica positiva, a diferencia del cristalino natural de los sujetos jóvenes; manifiestan, asimismo, que se suma a la aberración esférica positiva de la córnea.

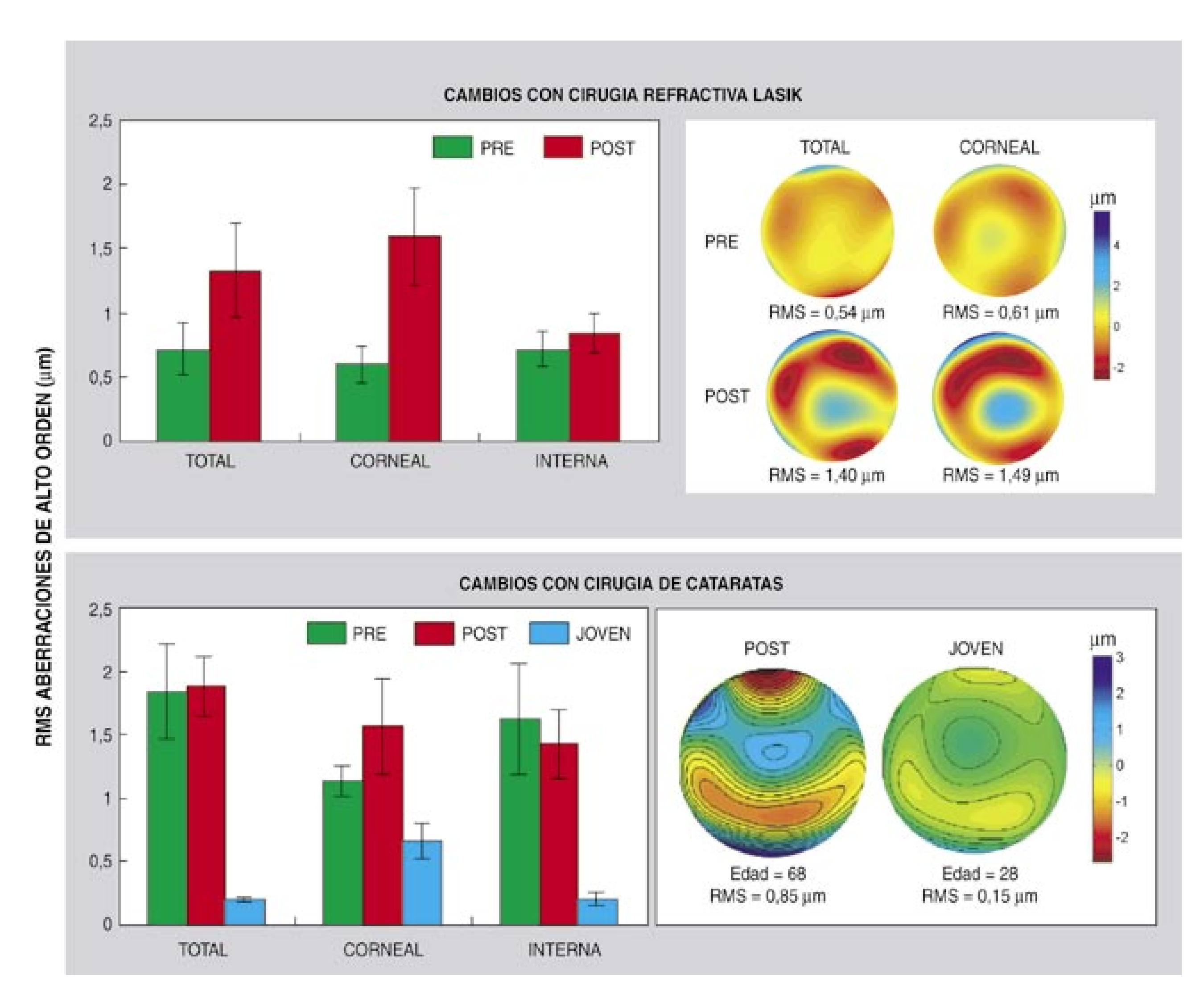
Las aberraciones de tercer orden y superior se incrementan también, probablemente a causa de desalineamientos y descentramientos de la lente intraocular. De esa investigación se desprende que, si bien la sustitución del cristalino con catarata por la lente intraocular proporciona una drástica mejora de la calidad de imagen retiniana, merced a la reducción de la difusión intraocular, no disminuyen las aberraciones.

Cabe, pues, esperar que el refinamiento del diseño de la lente intraocular y el continuo perfeccionamiento de los procedimientos quirúrgicos contri-



9. ERROR CUADRATICO MEDIO DEL FRENTE DE ONDA (RMS) para aberraciones de tercer orden y superior. Se estima en función de la acomodación (*a*), en función de la edad (*b*) y en función del error refractivo (*c*).





10. ERROR CUADRATICO MEDIO DEL FRENTE DE ONDA para aberraciones de tercer orden y superior corneales, totales e internas antes y después de cirugía refractiva LASIK para miopía (*panel superior*) y antes y después de cirugía de cataratas en comparación con un grupo de sujetos jóvenes normales (*panel inferior*). Los gráficos de barras muestran promedios entre varios sujetos (14 para el estudio de cirugía LASIK y 9 para el estudio de cirugía de cataratas), para 6,5 mm y 5 mm de pupila, respectivamente. La figura también recoge ejemplos típicos de mapas de aberraciones corneal y total antes y después de la cirugía refractiva LASIK (*panel superior*) y aberraciones totales en un paciente con una lente intraocular en comparación con un sujeto joven.

SUSANA MARCOS

buirán a mejorar la calidad óptica en pacientes operados de cataratas.

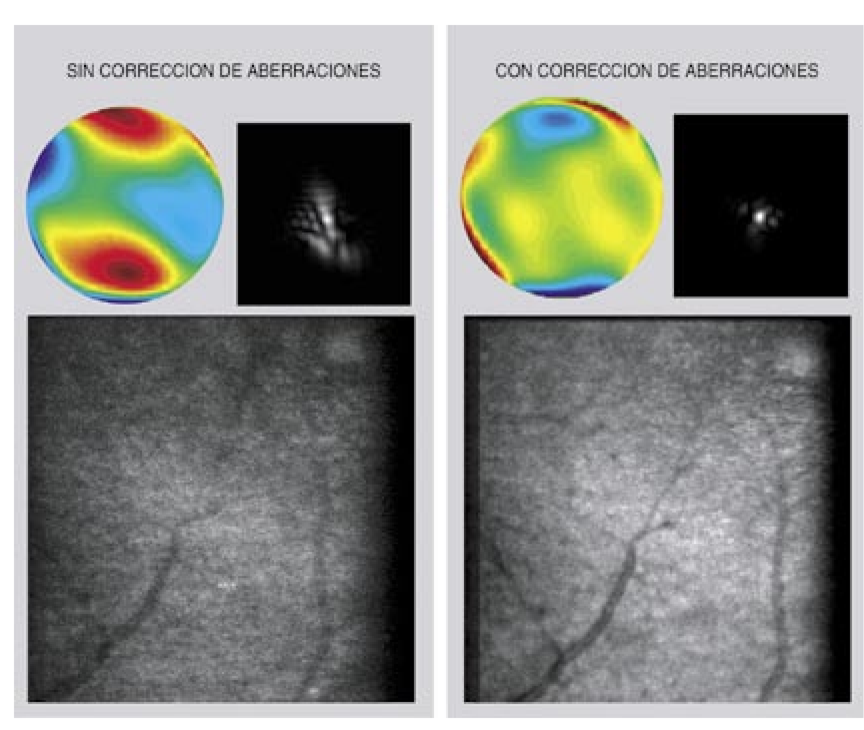
Mejora de la calidad de imagen retiniana del ojo

La medición precisa de las aberraciones del ojo abre la posibilidad de abordar su corrección. Una reducción de las aberraciones oculares mejoraría la calidad óptica de las imágenes que se proyectan sobre la retina y, posiblemente, la calidad perceptiva de dichas imágenes; se refinaría también la calidad óptica de las imágenes que se obtienen del fondo del ojo mediante oftalmoscopia: al aumentar el contraste y la resolución de las estructuras retinianas y, presumiblemente, elaborar un diagnóstico más ajustado y proceder a un seguimiento más fiel de las patologías retinianas.

Los nuevos equipos de cirugía refractiva LASIK se encuentran dotados de sistemas de aberrometría y prometen la corrección no sólo de las aberraciones de bajo orden, así el desenfoque y el astigmatismo, sino también de las aberraciones de alto orden.

Disponemos de pruebas experimentales de la posibilidad de corregir las aberraciones del ojo. Hasta la fecha se han utilizado sistemas de óptica adaptiva, una técnica desarrollada inicialmente para aplicaciones astronómicas. Emplea un espejo deformable para que adopte la forma inversa de la aberración de onda y, de ese modo, compensar el frente de onda que refleja, así como lentes de contacto especiales y láminas de fase individualizadas para el patrón de aberraciones particular de cada sujeto. El espejo deformable permite la compensación dinámica de las aberraciones. Las lentes de contacto se limitan a una compensación de la componente estática.

La figura 11 muestra el frente de onda de un sujeto antes y después



11. MEJORA DE LA CALIDAD DE LAS IMAGENES del fondo del ojo, obtenidas mediante un oftalmoscopio de barrido láser de gran aumento, con la compensación de las aberraciones oculares. En el panel izquierdo se ilustran las aberraciones de un sujeto, su correspondiente imagen de un punto y la imagen de la retina obtenida tras compensar únicamente el desenfoque. El panel derecho muestra el patrón de aberraciones del sujeto con cerca de un 70% de sus aberraciones corregidas, la imagen de un punto correspondiente y una imagen de la misma zona de la retina, obtenida mediante la compensación de las aberraciones de orden bajo y alto. Los vasos capilares de las fotografías revelan un importante incremento de la nitidez contraste y se visualizan estructuras retinianas irresolubles en la imagen convencional.

de la compensación de las aberraciones mediante una lámina de fase colocada en un plano conjugado a su pupila, y las PSF correspondientes. Las imágenes del fondo de ojo de este sujeto, obtenidas mediante un oftalmoscopio de barrido láser, construido en el Instituto Schepens de Investigación Ocular de Harvard, muestran una mejora del contraste y de la resolución de estructuras retinianas tras la corrección de aberraciones.

Conclusión

En definitiva, la nueva tecnología para la medida de la calidad óptica de la córnea y cristalino ha permitido en los últimos años avanzar en el conocimiento de los mecanismos de formación de imágenes del ojo en la retina y de los mecanismos de algunos procesos biológicos (como la acomodación y la presbicia, el envejecimiento del sistema ocular o el desarrollo de la miopía). En particular, esta tecnología está alcanzando una gran repercusión en la práctica clínica de la oftalmología y la optometría, dado su carácter objetivo y no-invasivo. La medida pormenorizada de las aberraciones oculares en pacientes operados de cirugía refractiva corneal para miopía ha revelado que los perfiles de tallado de la córnea por láser deben ser refinados para evitar, al tiempo que se corrigen los errores refractivos, inducir otras imperfecciones oculares. Asimismo, estas técnicas se han probado muy útiles en la evaluación de los resultados de la cirugía de cataratas con implante de lentes intraoculares, o en el diagnóstico de otras patologías oculares. La corrección de las aberraciones oculares abre la posibilidad no sólo de una mejora de la calidad visual de un individuo (una vez claras las relaciones entre la calidad óptica y la calidad visual perceptual), sino fundamentalmente de una mejora de la resolución y contraste de las imágenes del fondo del ojo. Previsiblemente la visualización de estructuras retinianas *in vivo*, hasta ahora sólo resolubles en muestras histológicas, permitirá avanzar en el diagnóstico precoz de ciertas patologías y una mejor monitorización de posibles terapias.

La autora

Susana Marcos, doctora en ciencias físicas por la Universidad de Salamanca, completó su formación en el Instituto Schepens de Investigación Ocular de la Universidad de Harvard. Actualmente es investigadora científica en el Instituto de Optica del CSIC en Madrid. Autora de numerosas publicaciones en las mejores revistas de su campo y conferenciante habitual en reconocidos congresos y prestigiosas instituciones extranjeras, ha recibido entre otros premios internacionales la Medalla Adolph Lomb, de la Sociedad Americana de Optica. Su grupo de investigación ha sido pionero en la evaluación objetiva de las propiedades ópticas del ojo y la retina, a través de novedosas técnicas de imagen, como sensores de onda y alta resolución del fondo del ojo.

Bibliografía complementaria

OPTICAL RESPONSE TO LASIK FOR MYOPIA FROM TOTAL AND CORNEAL ABERRATIONS. S. Marcos, B. Barbero, L. Llorente y J. Merayo Lloves en *Investigative Ophthalmology and Visual Science*, vol. 42, págs. 3349-56; 2001.

IMPERFECT OPTICS MAY BE THE EYE'S DEFENSE AGAINST CHROMATIC BLUR. J. S. McLellan, S. Marcos, P. M. Prieto y S. A. Burns en *Nature*, vol. 417, págs. 174-6; 2002.

CONTRIBUTION OF THE CORNEA AND INTERNAL SURFACES TO THE CHANGE OF OCULAR ABERRATIONS WITH AGE. P. Artal, E. Berrio, A. Guirao y P. Piers en *Journal of the Optical Society of America A-Optics Image Science and Vision*, vol. 19, págs. 137-43; 2002.

CONTRAST IMPROVEMENT FOR CONFOCAL RETINAL IMAGING USING PHASE CORRECTING PLATES. S. A. Burns, S. Marcos, A. E. Elsnser y S. Bará en *Optics Letters*, vol. 27, págs. 400-2; 2002.

SUSANA MARCOS

El método botánico de Cavanilles

En su obra culminó la botánica descriptiva de la Ilustración. Sus *Monadelphiae Classis dissertationes* y sus *Icones et descriptiones plantarum* contienen descripciones de centenares de especies de casi todo el mundo que hoy continúan unidas a su nombre en la Nomenclatura Botánica Internacional

José María López Piñero



1. Antonio José Cavanilles. Grabado sin firma.

JOSE MARIA LOPEZ PIÑERO

El segundo centenario de la muerte de Cavanilles ha sido una ocasión propicia para que los oportunistas reiteren la mitificación falseada y los errores que su biografía y su obra vienen padeciendo, sin tener en cuenta los numerosos trabajos de investigación sobre ambas en diferentes idiomas. Cerca de veinte veces se han reimpreso sus *Observaciones sobre la Historia Natural, Geografía, Agricultura, población y frutos del Reyno de Valencia* (1795-1797), obra menor en el conjunto de su producción, con el resultado de considerarlo como una especie de excursionista por tierras valencianas. Por el contrario, prácticamente se ignoran sus *Monadelphiae Classis dissertationes* (1785-1790) y sus *Icones et descriptiones plantarum* (1791-1801), a pesar de que contienen descripciones de centenares de especies de casi todo el mundo que hoy continúan unidas a su nombre en la Nomenclatura Botánica Internacional. Es muy significativo que la primera edición facsímil de las *Icones* fuera publicada en Holanda el año 1965, formando parte de la serie *Historia naturalis classica* y con estudio introductorio de una autoridad en taxonomía como F. A. Stafleu.

El presente artículo sólo intenta resumir los escasos trabajos de investigación sobre el método del gran botánico. Como el mejor sobre este tema publicado hasta ahora es un detenido análisis de sus *Dissertationes,* he procurado recientemente complementarlo con un estudio de la estructura y el contenido de las *Icones*.

Supuestos básicos del método de Cavanilles

Los estudios rigurosos han seleccionado dos textos sobre el método de Cavanilles. El primero procede de los *Principios elementales de la botánica* (1803), su principal obra teórica: "Puede reputarse cada una [planta] como un problema aislado, cuya solución pende de los principios fundamentales de la ciencia, reducidos a las partes de la fructificación y al orden del sistema adoptado". El segundo, de un discurso que pronunció el mismo año en el Jardín Botánico de Madrid: "Es tan regularmente exacta la ciencia de los vegetales, perfeccionada como hoy la tenemos, que ninguna de las naturales la lleva exceso en



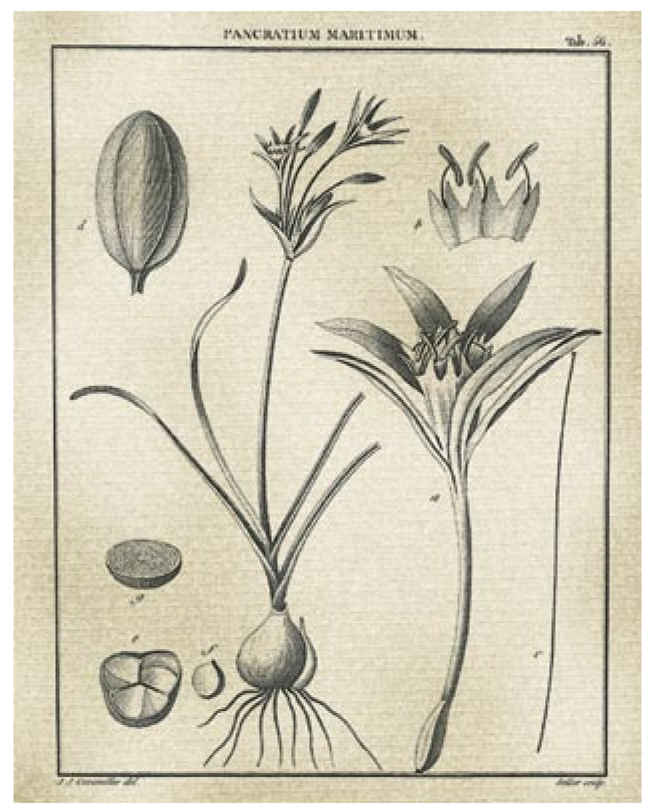
2. Trece especies del género botánico *Sida*. Lámina I de *Monadelphiae classis dissertationes decem* (1785-1790).

la exactitud. La organización de las plantas suministra datos que no se encuentran en el reino mineral. La uniformidad y constancia en reproducirse no da lugar a aquellos resultados opuestos que con frecuencia se observan en las operaciones químicas”.

Ambos textos reflejan de forma evidente aspectos fundamentales del método botánico de Cavanilles. Su sólida preparación en lógica y ciencias físico-matemáticas le condujo a tener como supuestos básicos el orden y la precisión. Por ello, se dedicó principalmente a la taxonomía, utilizando una modificación del sistema de Linné y manteniendo el fijismo del gran médico y naturalista sueco sobre las especies. A pesar de la influencia de la lógica no recurrió a razonamientos escolásticos, que tanto había despreciado en sus *Apuntamientos lógicos* (1780). Por el contrario, su *Denkstil* (estilo de pensamiento) concedió una función primordial a la experiencia. Sin hipérbole, puede afirmarse que en su obra culminó la botánica descriptiva de la Ilustración. En ello reside su relieve internacional, y no en que fue “el más importante naturalista español de todas las épocas”, como dicen sus hagiógrafos, olvidando o desconociendo otras figuras de tanta o mayor importancia: Nicolás Monardes, Francisco Hernández, José de Acosta, Félix de Azara, etc.

Cavanilles describió las plantas con un excepcional rigor morfológico, observando los más pequeños caracteres diferenciales con extraordinario detalle, no sólo a simple vista, sino también mediante las lentes y los microscopios de la época. En algunas ocasiones, lo hizo con el fin de describir diminutas formaciones, imprescindibles para aplicar debidamente sus criterios de clasificación: “Invisibles son sin el socorro de un fuerte microscopio las formas de los cotiledones, la dirección del rejo, la presencia o ausencia de la clara en multitud de semillas. Véase las de mis *Gesnerias,* apenas discernibles sin microscopio... Como restan infinitas plantas en que no se han hecho aún las debidas observaciones, como muchas de las hechas solamente se han podido practicar con el auxilio de fuertes microscopios... he creído deber llamar corola a la tela u órgano inmediato al germen o a los estambres”.

Por otra parte, se interesó por la indagación micrográfica de la textura íntima vegetal, aunque no estaba aún difundida la corrección técnica de la aberración cromática. En *Principios elementales de la botánica* (1803) describe, por ejemplo, el “tejido celular” en la raíz, el tallo y las hojas: “Las raíces gruesas se ven compuestas de muchas capas concéntricas, de las cuales la exterior, llamada epidermis, es sutil y rugosa, con multitud de agujeritos a donde van a parar los vasos por donde corre el jugo. Bajo esta membrana se halla el tejido celular, sustancia jugosa, compuesta de vejiguitas mezcladas con filamentos sutiles que se extienden en varias direcciones... En el tronco de las dicotiledóneas se halla el tejido celular, sustancia jugosa y ordinariamente verde, compuesta de granitos casi redondos, o bien sean vejiguitas mezcladas con filamentos muy sutiles que se prolongan en todas direcciones... Consta cada hoja



3. Azucena de mar, “lliri de marines” (*Pancratium maritimum* L.). Lámina 56 de *Icones et descriptiones plantarum* (1791-1801). Entre las numerosas plantas mediterráneas que Cavanilles estudió con detenimiento figura esta bella especie, una de las descritas por Juan Plaza durante el Renacimiento.





4. "Clavel de Indias" (*Tagetes tenuifolia* Cav.). Lámina 169 de *Icones et descriptiones plantarum* (1791-1801). Ejemplo de las veinte plantas americanas que había dado a conocer en Europa la gran expedición científica dirigida por Francisco Hernández (1571-1577) y cuya descripción postlinneana publicó Cavanilles. Ante la lógica falta de nombre europeo, Hernández la había denominado con el vocablo nahua "cempoaxóxitl".

de dos superficies o membranitas, una superior y otra inferior... quedando entre ellas un tejido de vasos, que se dividen y subdividen prodigiosamente... hay en este tejido multitud de vejiguitas".

Podría pensarse que se redujo a anotar los hallazgos ajenos y que los datos microscópicos de sus publicaciones son únicamente resultados de su erudición libresca. Nada más falso, ya que fue un microscopista práctico, como se manifiesta, entre muchos otros, en el texto siguiente: "Al examinar las cajitas y semillas de los helechos, he visto siempre ciertos cuerpecitos algo mayores que las semillas... Me he valido para esto del excelente microscopio de Dellabarre y de su lente número 4, que aumenta los objetos de manera que las cajitas parecen del tamaño de una lenteja".

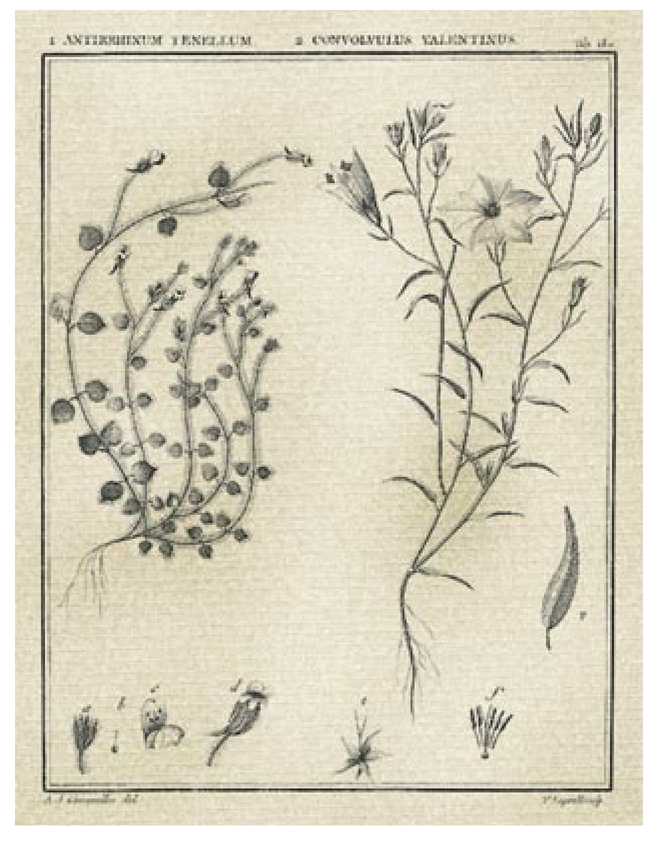
Doctrina fibrilar

No cabe duda de que figuró entre los numerosos adelantados del celularismo botánico, aunque, por supuesto, era seguidor de la doctrina fibrilar, doctrina vigente durante la Ilustración que consideraba la fibra como unidad elemental de la textura. Su inclinación a la física es coherente con los experimentos que realizó para cuantificar las características mecánicas de las fibras como, por ejemplo, su resistencia, que investigó mediante pesas. Esta vertiente fue la única avanzada de su fisiología vegetal, ya que eran muy tradicionales otros procedimientos que utilizó: las comparaciones de las siembras, los trasplantes, la observación del crecimiento al aire libre o en invernaderos, etc. Además, permaneció ajeno a las decisivas contribuciones del *Traité élémentaire de chimie* (1789), título fundacional de la química contemporánea, que Lavoisier publicó en París el mismo año en el que Cavanilles regresó definitivamente a Madrid. Ni siquiera llegó a interesarle la relación con Lavoisier que en el propio París había tenido Joseph Priestley (1774), cuando éste acababa de explicar la función del oxígeno en la fotosíntesis.

La destreza artística de Cavanilles, parecida a la que luego tendría Cajal, contribuyó a que lograse culminar la botánica descriptiva de la Ilustración en una línea principalmente morfológica, dibujando personalmente casi todas las láminas que ilustran sus publicaciones.

Función del herbario

En los estudios rigurosos se ha calificado a Cavanilles de "botánico de gabinete", afirmación que veremos confirmada al considerar las fuentes de las *Icones:* poco más de la cuarta parte de sus descripciones se basa en la observación directa de la naturaleza, una cifra similar en la de plantas de jardín y casi la mitad en herbarios. Linné había dado un gran impulso al uso de estos últimos, que pasaron a ser indispensables, no sólo para investigar, sino para demostrar después la objetividad de



5. Los endemismos botánicos valencianos *Chaenorrhinum tenellum* (Cav.) Lange y *Convolvulus valentinus* Cav. Lámina 180 de *Icones et descriptiones plantarum* (1791-1801) de Antonio José Cavanilles.

JOSE MARIA LOPEZ PIÑERO



6. *Piqueria trinervia* Cav. Especie botánica mexicana de un género que Cavanilles denominó en honor de Andrés Piquer, máxima figura de la medicina valenciana de la Ilustración. Lámina 235 de sus *Icones et descriptiones plantarum* (1791-1801).



7. *Dahlia coccinea* Cav.. Lámina 266 de *Icones et descriptiones plantarum* (1791-1801) de Antonio José Cavanilles.

JOSE MARIA LOPEZ PIÑERO

las descripciones. Cavanilles fue seguidor suyo también en este aspecto metodológico, usando los herbarios de forma equivalente a la actual desde casi todos los puntos de vista. Dada la importancia que tuvieron para su obra, los defendió repetidas veces, sobre todo frente a los que destacaban la importancia de complementarlos con la observación de plantas vivas.

En su *Colección de papeles sobre controversias botánicas* (1796), donde recogió el lamentable enfrentamiento con Casimiro Gómez Ortega y su discípulo Hipólito Ruiz, afirmó: "Por más activo que sea un botánico y aunque consuma su vida en viajar, jamás podrá ver más de una porción de vegetales, comparados con los innumerables que existen; y al fin, para instruirse a fondo, recurrirá al huerto seco o herbarios". Sin embargo, se vio obligado a citar un párrafo del prólogo de la *Quinología* (1792) de Ruiz, primera publicación con materiales de la expedición al Perú y Chile: "Trepando por montes casi inaccesibles, y nunca señalados de huella humana, y por árboles encumbrados, cuyas ramas se me desgajaban bajo los pies, para tener la satisfacción de poder presentar al público mis descripciones hechas a la intemperie, entre aquellas asperezas y con la fatiga y peligro que conoce cualquier botánico experimentado, que sabe apreciar y distinguir estos trabajos de los que se hacen a la sombra y comodidad de un gabinete". Le hizo, además, la concesión siguiente: "Yo sé por experiencia cuán difícil es perfeccionar la obra en los montes; y si he logrado hacerlo en los del Reino de Valencia, menos ricos sin comparación en vegetales que el suelo americano, debe atribuirse al corto número de plantas nuevas que he descubierto, no pasando de 200 en tres años de viajes... Y por esto infiero con mucha probabilidad lo que habrá sucedido al señor Ruiz en aquel país rico... no de una clase o familia solamente, sino de todas las conocidas en el reino vegetal". Al margen de la penosa polémica, Cavanilles tenía muy claras las desventajas de describir especies botánicas basándose exclusivamente en el estudio de herbarios. Por ello, en las *Icones,* fue indicando con minuciosidad las limitaciones de los ejemplares incompletos.

Cuando pudo observar las plantas directamente en su hábitat natural, realizó aportaciones de interés a la disciplina que desde Ernst Haeckel (1866) se llama ecología. Como vamos a ver, en 201 especies de este grupo descritas en las *Icones* incluyó breves listas sobre la vegetación de la zona y en siete de ellas añadió estudios ecológicos más extensos. En las *Observaciones* expuso ampliados los correspondientes al territorio valenciano.

Relación con Lamarck

Cavanilles mantuvo también el fijismo de Linné sobre las especies. Este supuesto básico basta para explicar su actitud ante las hipótesis evolucionistas de Buffon y de Lamarck. Los que pretenden que ignoró la de Lamarck porque éste la expuso por vez primera en 1799, cuando Cavanilles ya residía definitivamente en Madrid, desconocen las vías a través de las cuales funcionaba en la

época la información científica y no tienen en cuenta algo tan obvio como las fechas hasta las que se mantuvo la correspondencia entre ambos.

Tampoco sirve recurrir a la creciente enemistad que los fue distanciando. A este respecto es fundamental como fuente la ardiente defensa de Linné frente a determinadas críticas de Lamarck que figura casi al comienzo de las *Icones,* en la que llega a decir que apenas entiende “cuál es la razón que ha conducido al Señor de Lamarck a introducir tanta injuria en la memoria de Linné”. Está todavía más claro que el motivo de su postura ante la hipótesis evolucionista de Buffon fue su adhesión al fijismo, de acuerdo con los dos supuestos básicos que antes hemos anotado: el orden y la precisión. Repetimos que con Cavanilles culminó el nivel inicial del saber botánico, integrado por la descripción. Para alcanzar un nivel explicativo era necesaria la teoría evolucionista, fundamento de los sistemas taxonómicos a partir del darwinismo.

Las *Dissertationes*

Cavanilles inició su dedicación a la botánica con la serie, antes citada, de diez *Dissertationes* sobre la “clase” del sistema de Linné *Monadelphiae.* Las ocho primeras las publicó en París (1785-1789) y las dos últimas, en Madrid (1790). Fueron una especie de fascículos del libro en tres volúmenes *Monadelphiae Classis Dissertationes Decem,* editado en Madrid en el mismo año e imprenta que las dos últimas, con un prefacio de gran interés para conocer sus cambiantes criterios taxonómicos. En ellas utilizó plantas de terrenos más o menos cercanos a París y de los que observó durante sus viajes junto al duque del Infantado, aunque fueron más numerosas las que crecían en el *Jardin du Roy* y el de Trianon, en los de Jacques Martin Cels, del propio duque, etc. Muchos ejemplares del *Jardin du Roy* procedían de expediciones, entre ellas, las dirigidas por René L. Desfontaines al norte de Africa (1783-85) y por Hipólito Ruiz y José Pavón a Perú y Chile (1777-1788), desde donde los remitía Joseph Dombey. También aprovechó las plantas que le enviaron Gómez Ortega y Antonio Palau desde Madrid y Cándido María Trigueros desde Carmona, así como los ricos herbarios y colecciones de pinturas y dibujos que tenían los Jussieu y Lamarck.

Todo ello explica que figurasen en ellas numerosas especies exóticas, como, las asiáticas denominadas en hindi “bhiunli” (*Sida humilis* Cav.), y “bariara” (*S. carpinifolia* L.) o las americanas llamadas “thipon” en huaxteca (*S. pyramidata* Cav.), “zakmizbil” en maya (*S. cordifolia* L.) y “axocatzin” en náhuatl (*S. rhombifolia* L.). También, que el género *Palaua* Cav. proceda de los Andes y *Anoda* Cav., de los territorios desde México hasta Chile y Argentina.

Estructura y contenido de las *Icones*

Las *Icones* incluyen la descripción de 712 especies, que Cavanilles estudió observando plantas silvestres, cultiva-



8. *Eucaliptus rostratus* Cav., especie botánica australiana. Lámina 342 de *Icones et descriptiones plantarum* (1791-1801) de Antonio José Cavanilles.



JOSE MARIA LOPEZ PIÑERO

9. *Stevia salicifolia* Cav. Especie botánica mexicana de un género que Cavanilles denominó en honor de Pedro Jaime Esteve. Lámina 354 de sus *Icones et descriptiones plantarum* (1791-1801).



10. *Guioa leniscifolia* Cav. Especie botánica de las islas Vavao de un género que Cavanilles denominó en honor de José Guío, uno de los artistas de la expedición Malaspina. Lámina 373 de sus *Icones et descriptiones plantarum* (1791-1801).



das en jardines o herbarios, de acuerdo con la siguiente distribución:

VOLUMENES	SILVESTRES	JARDINES	HERBARIOS	TOTALES
I	47	58	4	109
II	91	19	-	110
III	41	68	-	109
IV	4	36	74	114
V	-	11	119	130
VI	4	12	124	140
Totales	187	204	321	712
Porcentajes	26,26 %	28,65 %	45,08 %	

En algunas especies, sobre todo en las que estudió a través de herbarios, Cavanilles anota *ignoratur patria, patriam ignoro* y otras indicaciones equivalentes. Esta es la procedencia de los ejemplares:

PROCEDENCIA DE LOS EJEMPLARES DE LAS ESPECIES (EN LAS QUE CONSTA)

VOLUMENES	I	II	III	IV	V	VI	TOTALES	PORCENTAJES
Territorio valenciano	6	90	38	6	-	2	142	21,54 %
Otras zonas ibéricas	49	1	8	-	1	7	66	10,01 %
Islas Canarias	1	-	1	-	-	-	2	0,30 %
Africa	4	-	-	-	-	4	8	1,21 %
Asia	3	1	-	1	-	-	5	0,75 %
América tropical	3	-	-	4	-	1	8	1,21 %
Nueva España	37	16	49	32	36	-	170	25,79 %
Virreinato del Perú	9	1	5	20	50	30	145	17,45 %
V. Nueva Granada	-	-	-	2	-	-	2	0,30 %
V. Río de la Plata	1	-	-	5	26	12	44	6,52 %
Islas del Pacífico*	-	-	-	3	3	5	11	1,66 %
Australia	-	-	-	32	4	28	64	9 ,71 %
Islas Filipinas	-	-	-	8	14	-	22	3,33 %
Totales	57	93	41	7	-	3	659	

*Vavao (6) y Marianas (5)

Tras sus viajes por el territorio valenciano (1791-1793), expuso principalmente sus plantas en los volúmenes II y III. Las de Nueva España se reparten sin grandes diferencias desde el I hasta el V, porque primero utilizó los materiales de la expedición dirigida por Martín de Sessé (1788-1802) y luego los de la circunnavegación dirigida por Alejandro Malaspina (1789-1794). En los tres últimos se concentran las especies de Perú y Chile, que motivaron tantas polémicas, y de forma todavía más acusada las que conoció gracias a la citada circunnavegación. A pesar de su estrecha amistad con Mutis, solamente estudió dos de Nueva Granada.

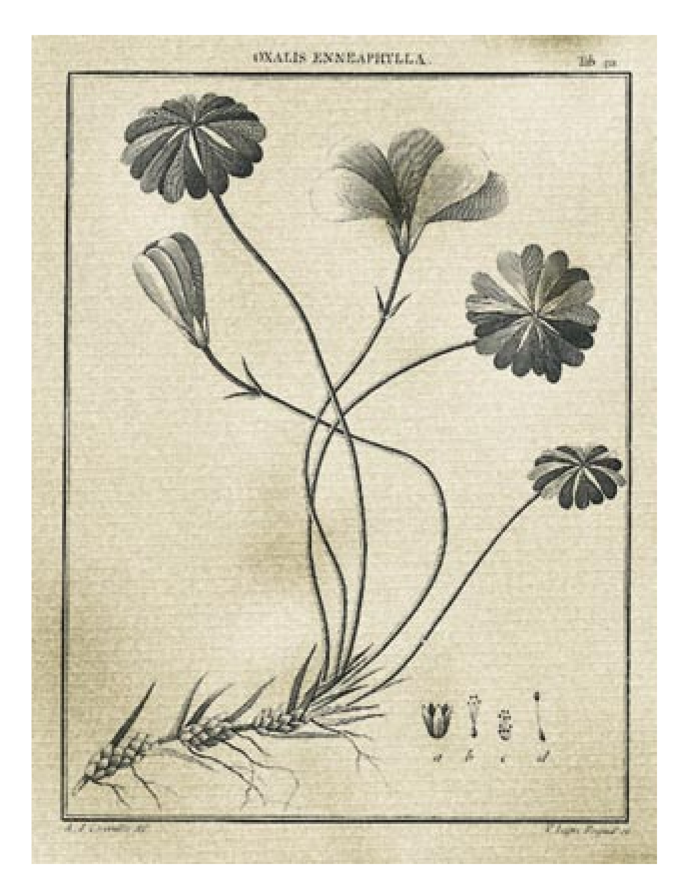
JARDINES*

VOLUMENES	I	II	III	IV	V	VI	TOTALES	PORCENTAJES
D. del Infantado (París)	7	-	-	-	-	-	7	3,46 %
Trianon	1	-	-	-	-	-	1	0,49 %
Botánico de París	-	-	1	3	-	-	4	1,98 %
Boutin (París)	-	-	-	-	1	-	1	0,49 %
Botánico de Madrid	45	18	65	34	10	12	184	91,09 %
La Priora de Madrid	2	-	1	-	-	-	3	1,48 %
Puzol	-	2	-	-	-	-	2	0,99 %
Totales	55	20	67	37	11	12	202	

*En las especies que consta el jardín (Cavanilles anota varios en algunas)

11. *Pomaria glandulosa* Cav. Especie botánica mexicana de un género que Cavanilles denominó en honor de Jaime Honorato Pomar. Lámina 402 de sus *Icones et descriptiones plantarum* (1791-1801).





12. *Oxalis enneaphylla* Cav., especie botánica de las islas Malvinas. Lámina 411 de *Icones et descriptiones plantarum* (1791-1801) de Antonio José Cavanilles.

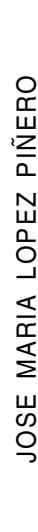
En el volumen I todavía figuran especies que había estudiado en los jardines del duque del Infantado y en el de Trianon, pero a partir del II predomina de modo terminante el Botánico de Madrid; el de La Priora era el de los boticarios de la Corte. El único valenciano que cita, sólo en dos especies, es el de Puzol, lo que se explica principalmente porque lo había organizado Vicente Alfonso Lorente, discípulo de su rival Gómez Ortega.

Es impresionante el número de especies que describió basándose en el herbario de Luis Neé, integrante de la circunnavegación dirigida por Malaspina dentro del grupo de naturalistas encabezado por el guatemalteco Antonio Pineda. Un claro ejemplo de la diferencia entre los "botánicos de campo" y los "de gabinete", que aprovechan cómodamente sus materiales, es la muerte de Pineda en Filipinas.

HERBARIOS*

VOLUMENES	I	II	III	IV	V	VI	TOTALES	PORCENTAJES
Neé	-	-	-	71	119	120	310	97,18 %
Broussonet	-	-	-	-	-	4	4	1,25 %
Jussieu	1	-	-	-	-	-	1	0,31 %
Mutis	-	-	-	1	-	-	1	0,31 %
Palau	1	-	-	-	-	-	1	0,31 %
Pavón	-	-	-	1	-	-	1	0,31 %
Thouin	1	-	-	-	-	-	1	0,31 %
Totales	3	-	-	73	119	124	319	

*En las especies de que consta el herbario (Cavanilles anota varios en algunas)



El interés de Cavanilles por la vegetación es muy claro en las descripciones de las plantas silvestres que estudió directamente. Aprovechó para observarla sus viajes a la localidad toledana de Méntrida —donde tenía una finca el duque del Infantado— y, sobre todo, los que realizó por el territorio valenciano, publicando las siguientes síntesis en las *Icones*:

- "Mentridae descriptio naturalis".
- "Castulonis descriptio naturalis".
- "Observationes" [sobre el algarrobo en el territorio valenciano].
- "De Saguntino monte".
- "Albaydae Vallis descriptio physica".
- "De Montibus Enguerinis".
- "De Orospeda, Aytana et Idubeda" [sierras de Mariola y Aitana y cordillera Ibérica].

Además, incluyó breves notas sobre el tema en la descripción de 201 especies:

NOTAS SOBRE VEGETACION

VOLUMENES	I	II	III	IV	V	VI	TOTALES	PORCENTAJES
Territorio valenciano	6	88	38	6	-	2	140	69,65 %
Madrid y cercanías	34	3	2	-	-	1	40	19,90 %
Aranjuez	7	-	-	-	-	-	7	3,48 %
Méntrida	6	1	-	-	-	-	7	3,48 %
La Mancha	1	1	-	-	-	-	2	0,99 %
Castilla-León	1	-	-	1	-	-	2	0,99 %
Andalucía	-	-	1	-	-	-	1	0,49 %
Aragón	-	-	1	-	-	-	1	0,49 %
Cantabria	1	-	-	-	-	-	1	0,49 %
Totales	57	93	41	7	-	3	201	

Apenas hace falta decir que, tras su regreso definitivo a Madrid, Cavanilles recorrió los alrededores, tanto "suburbios y márgenes de cultivo" como "terrenos áridos", fue a los valles de los ríos Manzanares y Jarama, subió a los altos de San Bernardino y al cerro Negro, observó "Sitios Reales" como la Casa de Campo, el Retiro y Aranjuez.

El elevado número de especies descritas en las *Icones* obliga a que un artículo como el presente se limite a meros ejemplos.

Cavanilles describió por vez primera endemismos del territorio valenciano como la escrofulariácea *Chaenorrhinum tenellum* (Cav.) Lange —llamándola "Antirrhinum tenellum"—, la convolvulácea *Convolvulus valentinus* Cav. y la dipsacácea *Scabiosa saxatilis* Cav. También, otras plantas que crecen en el mismo, como las orquídeas "abelletes grogues" (*Ophrys lutea* Cav) y "abelles" (*Ophrys scopolax* Cav), el rebollo (*Quercus valentina* Cav.) y la "cresolera" (*Phlomis crinita* Cav.). Estudió, además, con detenimiento varias conocidas, entre ellas, las ocho siguientes:

- el "cospí marí" (*Pseudorlaya pumila*), en la desembocadura del río Júcar.
- la "paternostrera" (*Withania frutescens* [L.] Pauquy), en Orihuela.
- la "margallonera" (*Chamaerops humilis* L.), única palma nativa de Europa, en el desierto de las Palmas, Jávea y Denia.
- la palmera datilera (*Phoenix dactylifera* L.), en el palmeral de Elche.
- la zamarrilla (*Teucrium polium* L.), en las colinas de Alcoy, Moixent y Enguera.



13. *Carmona heterophylla* Cav. Especie botánica de las islas Marianas de un género que Cavanilles denominó en honor del artista Bruno Salvador Carmona. Lámina 438 de sus *Icones et descriptiones plantarum* (1791-1801).

• el "pinell", pinillo de oro o corazón de piedra (*Hypericum ericoides* L.), en las grietas de las colinas de Sagunto, Xàtiva, Cullera, Enguera y Bocairente.

• la anagálide de pantano (*Anagallis tenella* [L.] L. fil.), en Buñol, Cortés de Pallás y Sumacárcer.

• el quejigo (*Quercus faginea* Lam.), en Montesa.

Párrafo aparte merecen el bello "lliri de marines" o azucena de mar (*Pancratium maritimum* L.) y el *Cynanchum acutum* L., porque completó las descripciones que durante el Renacimiento había hecho Juan Plaza cuando ocupaba en la Universidad de Valencia la cátedra de medicamentos simples (o "herbes") —una de las primeras de Europa— y que, como anota Cavanilles, difundió el neerlandés Charles de l'Escluse (Clusius) a través de su famosa obra *Rariorum aliquot stirpium per Hispanias observatarum Historia* (1576).

En la Casa de Campo madrileña y en el arroyo de Cantarranas comprobó que era raro *Anthyllis lotoides* L., endemismo de las zonas sin cultivar de la península Ibérica, y que abundaba en Méntrida. Otro endemismo ibérico que describió fue *Ortegia hispanica* L., indicando que ya la habían estudiado Clusius y Löffling, lo que supuso reconocer el género que Gómez Ortega había denominado en honor de su tío, el boticario José Ortega. Vio la *Avena loefflingiana* L. en los "terrenos áridos" de los alrededores de Madrid, pero advierte que, según Linné, es nativa del cabo de Buena Esperanza.

De su amplia contribución al conocimiento de la flora americana, anotaremos, en primer término, la formulación del género *Dahlia* y la descripción de la especie *Dahlia pinnata* Cav., de la que proceden variedades muy importantes hoy en jardinería. En segundo, la descripción postlinneana de veinte especies botánicas que había estudiado Francisco Hernández, en el curso de la primera gran expedición científica a América, que recorrió Nueva España durante siete años (1571-77).

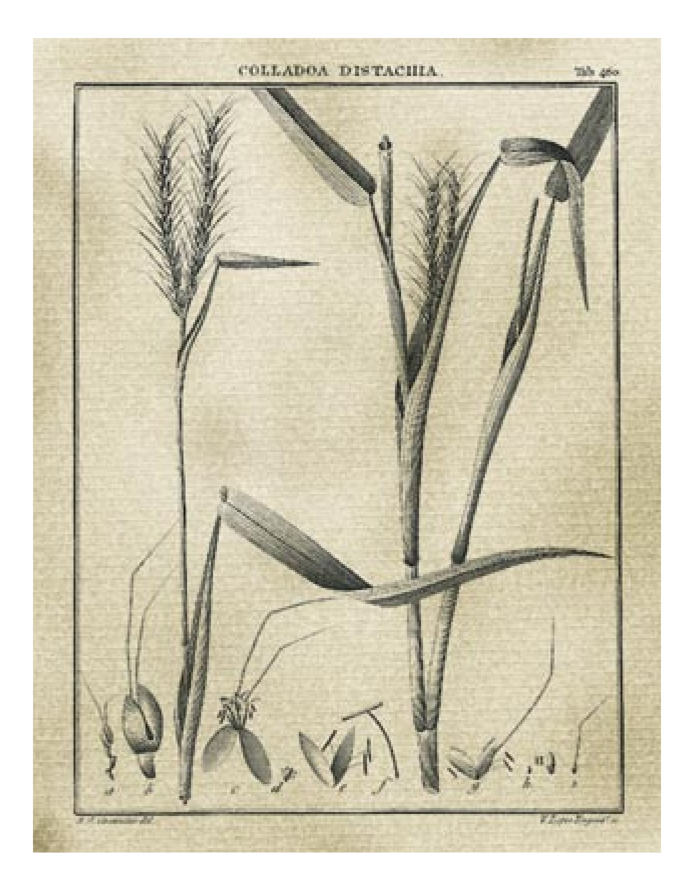
Recordemos que la contribución de Hernández fue gigantesca, ya que su *Historia plantarum Novae Hispaniae* contiene descripciones de casi tres mil especies vegetales, casi todas basadas en el estudio directo sometido a una pauta metódica rigurosa. Las ediciones impresas de los materiales hernandinos durante la primera mitad del siglo XVII tuvieron un importante peso en el punto de partida de la botánica moderna, a través de figuras como Joseph Pitton de Tournefort, John Ray, Linné y las expediciones científicas españolas de la Ilustración. Linné, que acordó con el gobierno español (1751) una misión científica destinada al estudio de la flora americana, enviando a Pehr Löffling, su discípulo predilecto, era un ferviente admirador de Hernández, como se manifiesta en su correspondencia con Löffling y en varios de sus trabajos monográficos, especialmente en el que publicó en las *Handligar* de la *Svenska Vetenskapakademie* (1755) sobre el "atzoatl" hernandino, que identificó como *Mirabilis longiflora.*

Las veinte plantas hernandinas cuyas descripciones postlinneanas publicó Cavanilles corresponden a los siguientes capítulos de la *Historia plantarum Novae Hispaniae* (edición madrileña de 1790 y mexicana de 1959):

CAPITULO	NOMBRE	ESPECIE
I, 24	Acocoxóchitl	*Dahlia coccinea* Cav.
I, 37	Apancholoa	*Cuphea aequipetala* Cav.
I, 41	Acxóyatl de Ocopetlayuca	*Asclepias linaria* Cav.
I, 43	Primer tlayacxóyatl	*Asclepias linaria* Cav.
XVI, 39	Tezonpatli	*Asclepias linaria* Cav.
I, 49	Segundo acxóyatic	*Galinsoga parviflora* Cav.
I, 62	Primer tonalxíhuitl	*Stevia saliciflora* Cav.
I, 112	Segundo chimalácatl	*Verbesina crocata* (Cav.) Less.
V, 161	Zacanahuitéputz	*Verbesina crocata* (Cav.) Less.
I, 115	Acatzannaícxitl	*Pellaea ternifolia* (Cav.) Ling.
II, 95	Coapatli origanoide	*Lippia umbellata* Cav.
III, 40	Itlanexillo	*Pellaea cordata* Cav.
IV, 112	Coapatli asfodelino	*Tradescantia crassiflora* Cav.
IV, 179	Cempoalxóchitl	*Tagetes peduncularis* Cav.
		Tagetes tenuiflora Cav.
V, 39	Tochcuitlaquílitl	*Salvia chamaedriodes* Cav.
V, 184	Chilpan	*Penstemon campanulatus* Cav.
VI, 78	Capolin	*Prunus capuli* Cav.
VI, 126	Curúhcume	*Tagetes lucida* Cav.
VI, 152	Yyauhtli	*Tagetes lucida* Cav.
VII, 17	Tercera cuerapocirangua	*Brickelia squarrosa* (Cav.) Wild.
VII, 37	Quinto totoncapatli	*Thryallis glauca* (Cav.) Kuntze
XIII, 32	Nacazpipilolli	*Exogonium bracteatum* (Cav.) Choisy
XVII, 4	Tlacoxóchitl anenecuilcense	*Bouvardia terniflora* (Cav.) Schlecht.

Cavanilles denominó el género *Stevia* en honor de Pedro Jaime Esteve, antecesor de Plaza en la cátedra valenciana de "herbes" y autor de un *Diccionario de las yerbas y plantas medicinales que se hallan en el Reino de Valencia* que circuló, como tantos otros libros de la época, en copias manuscritas. Dedicó *Pomaria* a Jaime Honorato Pomar, sucesor de Plaza y consejero botánico de Felipe II, que

JOSE MARIA LOPEZ PIÑERO



14. *Colladoa distachia* Cav. Gramínea de las islas Filipinas de un género que Cavanilles denominó en honor del célebre anatomista valenciano del Renacimiento Luis Collado. Lámina 460 de sus *Icones et descriptiones plantarum* (1791-1801).

JOSE MARIA LOPEZ PIÑERO

le donó el espléndido atlas hoy llamado Códice Pomar (*ca.* 1590), y *Calvoa* al cirujano Juan Calvo, que había dedicado también gran atención a las plantas medicinales en la Valencia renacentista. Incluyó el "tubal" o "toubai" (*Piqueria trinervia* Cav.) en un género con cuyo nombre homenajeó igualmente al turolense Andrés Piquer, máxima figura de la medicina valenciana de la Ilustración. Otros dos ejemplos significativos de las plantas americanas que contienen las *Icones* son:

- la malva de hojas estrechas (*Sphaeralcea angustifolia* [Cav.] G. Don.), que describió bajo el término *Malva angustifolia.* Dice que "honestissimus vir Antonius Palau" le había enviado en 1786 a París un ejemplar y que, en su *Tertia dissertatio,* le había prometido la descripción completa que ahora hace.
- la *Galinsoga parviflora* Cav. cita a Ruiz y Pavón, admitiendo el género *Galinsoga,* que éstos habían formulado. Otra prueba de que la polémica tuvo sus límites.

A este respecto, son muy significativas las dos comparaciones que ofrece con las *Novarum aut rariorum plantarum ... Descriptionum decades* (1797-1800), de Gómez Ortega, además de citarlas en seis especies: cinco mexicanas y una chilena. Antes de la primera comparación, se refiere a las plantas que habían llegado al Jardín Botánico madrileño de los "Reinos mexicano y peruano", así como a los apuntes de Neé sobre el lugar donde crecían, añadiendo una detallada lista de 82 especies.

Las plantas de Australia, a la que Cavanilles llama "Nova-Hollandia", pueden ejemplificarse en las descripciones de las especies de *Eucaliptus*, *Embothrium* y *Protea,* todas ellas basadas *in laudato herbario* de Neé:

- *Eucaliptus corymbosus, E. platypodus, E. rostratus*, *E. obliquus, E. salicifolius y E. racemosus*, las tres últimas sin lámina y con una descripción muy sucinta. Sólo cita a l'Heritier y a Lamarck.
- *Embothrium herbaceum, E. linearifolium, E. cytysoides, E. genianthum* y *E. spathulatum.* En la descripción de la primera dice: "Los autores de la flora peruana y chilena han ilustrado este género de Forster con nuevas especies... a las cuales añado ésta y las siguientes". Se refiere a *Florae Peruvianae et Chilensis Prodromus* (1794), de Ruiz y Pavón, ya que su *Flora Peruviana et Chilensis* inició su publicación después del volumen IV de las *Icones.* Actualmente el género es *Embothrium* Forster et Forster, de las proteáceas. Dentro de esta serie, Cavanilles incluye *E. ferrugineum,* una especie de Chile.
- *Protea tridactylides, P. acufera, P. pulchella y P. dichotoma.* En la tercera, cita la descripción que había publicado el médico alemán Heinrich Adoph Schrader, director del Jardín Botánico de Göttingen, en el volumen primero de su *Sertum Hannoveranum, seu plantae rariores, quae in hortis regiis Hannoverae vicinis coluntur* (1795).

Guioa leniscifolia Cav. es una especie de las islas Vavao de un género que Cavanilles denominó en honor de José Guío, "disecador y dibujante" de la expedición Malaspina. También con el nombre genérico de *Carmona heterophylla* Cav., especie de las islas Marianas, homenajeó a otro artista: Bruno Salvador Carmona. Ambos términos se mantienen en la Nomenclatura Botánica Internacional.

De las veintidós procedentes de las islas Filipinas descritas en las *Icones,* citaremos como ejemplos tres gramíneas: *Aristida murina, Aristida rigida* y *Colladoa distachia.* Sobre la base de esta última formuló un género obviamente dedicado a Luis Collado, el célebre anatomista valenciano del Renacimiento.

El autor

José María López Piñero, nacido en Mula (Murcia) el año 1933, es catedrático jubilado de Historia de la Medicina. Fundó el hoy llamado Instituto de Historia de la Ciencia y Documentación López Piñero (CSIC-Universidad de Valencia), único centro español en su género.

Bibliografía complementaria

A. J. CAVANILLES. F. A. Stafleu, R. S. Cowan en *Taxonomic Literature. A Selective Guide to Botanical Publications and Collections with Dates, Commentaries and Types*, 2ª ed., vol. I: A-G. Scheltema & Holkema, págs. 471-474; Utrecht, Bohn, 1976.

RELECTURA DE LAS *DISSERTACIONES* DE CAVANILLES. J. M. Valderas en *Collectanea Botánica*, n.º 20, págs. 183-238; 1991.

BIBLIOGRAFÍA DE ANTONIO JOSÉ CAVANILLES (1745-1804) Y DE LOS ESTUDIOS SOBRE SU VIDA Y SU OBRA. 3ª ed. actualizada. J. M. López Piñero, M. L. López Terrada en la edición facsímil digitalizada de *Icones et descriptiones plantarum...*, Real Sociedad Económica de Amigos del País; Valencia, 2004.

CURIOSIDADES DE LA FÍSICA

Norbert Treitz

Juegos con el centro de masas

Decía Einstein que la física describía la naturaleza con la máxima sencillez posible, pero ni un gramo más. Arboles y edificios, dobladas sus trayectorias por fuerzas aparentes, vuelan a nuestro alrededor cuando los observamos desde un coche que traza una curva. Tomar ese punto de vista no es incorrecto; tampoco es inteligente. Una descripción mucho más simple se consigue en un sistema inercial, un sistema de referencia donde los objetos no obedecen a fuerzas aparentes, sino a fuerzas internas del sistema y, ante todo, a la propia inercia. En este caso, la misma conservación del movimiento (la tercera de las famosas leyes de los "Principia mathematica" de Newton) nos dice en términos generales qué *no* puede suceder.

Si se tiene un sistema inercial, cada sistema de referencia que se mueva con respecto al mismo a velocidad constante será también inercial. (Constante en dirección y magnitud, se sobreentiende; ¡no se refiere a sistemas en rotación!) Más aún: todos estos sistemas son equivalentes. No se puede distinguir uno de otro con experimentos mecánicos limitados a cada uno de esos sistemas. En concreto, no hay ningún sistema de referencia al que se pueda calificar de absolutamente quieto. Ese es el principio de relatividad. Pero con la teoría de la relatividad apenas si tiene en común más que el nombre. Ese principio regía ya en la mecánica clásica (o "no relativista") de la que se habla aquí.

La validez del principio se aprecia al servir una bebida en un AVE en marcha; o, simplemente, si se tiene en cuenta que la Tierra rota con una considerable velocidad tangencial instantánea y además se traslada aún más rápidamente alrededor del Sol y del centro de la galaxia. No hay que preocuparse por dichas velocidades, que no se pueden medir por medios mecánicos.

Para cada interacción entre dos objetos, hay, entre todos los sistemas de referencia equivalentes, uno de interés particular: aquel que permite una descripción directa de lo que sucede. Nos referimos al sistema en el cual el centro de masas común a tales objetos se halla en reposo. En dicho sistema, la suma de las energías cinéticas es menor que en cualquier otro.

De aquí en adelante trataremos de los choques, de las interacciones en las que el momento lineal se transfiere de un objeto a otro de repente (esto es, en un tiempo muy breve y a lo largo de cortísimos recorridos): en los largos tiempos que preceden y siguen al choque los vectores de momento permanecen constantes. Un choque se dice que es elástico si la suma de sus energías cinéticas es igual antes y después —¡no necesariamente durante el trasvase de momento!—; será totalmente inelástico si al final

1. Sorprende lo alto que bota una pelota sobre otra mayor si se dejan caer ambas a la vez.





SPEKTRUM DER WISSENSCHAFT

ambos objetos se quedan quietos uno respecto al otro. Entre uno y otro extremo se dan la mayoría de los casos reales, demasiado complicados para nosotros.

Se busca donante: Dos pelotas *A* y *B* chocan elásticamente. Durante el choque, *A* proporciona un impulso ***dp*** a *B*; significa que el momento de *A* ha disminuido en una cantidad ***dp***, la misma en que ha crecido el de *B*. Supondremos algo más: que no hemos observado ningún intercambio de momento angular. Pero, ¿quién aporta la energía cinética? Si la pelota *A* está inicialmente quieta y luego no, será receptora de energía. Si es la pelota *B* la que está inicialmente quieta, nos hallaremos en la situación inversa: *A* será la donante. Debido al principio de relatividad, estas dos descripciones explican el mismo proceso en dos sistemas de referencia inerciales.

Podemos optar por un camino más fácil y utilizar el centro de masas común a ambas pelotas. Antes del choque, las pelotas tienen momentos de sentido opuesto e igual magnitud; después también, pero sus direcciones (opuestas) dependen de los detalles del contacto —los malos jugadores de billar se lo dejan al azar—. Gracias a esta simetría, en el sistema del centro de masas sólo se deben distinguir tres casos, según sean las magnitudes de ambos momentos tras el choque:

- (a) mayores que antes,
- (b) iguales que antes o
- (c) menores que antes.

El caso (c) aparece sobre todo cuando hay fricción, o sea, transformación de energía a formas que sólo se percibirán macroscópicamente como un aumento de temperatura. El caso (a) se da cuando, en el choque mismo, el sistema recibe energía de otras fuentes. Los buscapiés constituyen un buen ejemplo.

El caso (b) es el choque elástico. Debido a que los momentos tienen el mismo módulo en el sistema del centro de masas, no sólo la suma de las energías cinéticas es la misma antes y después, sino que cada pelota recupera su energía cinética (durante el contacto puede estar total o parcialmente almacenada como energía de deformación). En el caso de una fuerza gravitatoria, la energía cinética durante el contacto es mayor que antes y después del mismo.

Cuando se trata de choques elásticos entre dos objetos, en el sistema del centro de masas se transmite momento, aunque no energía: un resultado al que hay que acostumbrarse. Cuando se pasa de un sistema de referencia a otro, no sólo cambian momentos y energías, sino también los papeles de donante y receptor de energía.

Choque frontal: Dos agentes de tráfico discuten sobre el grado de gravedad de un accidente: ¿es mayor cuando choca un par de coches de igual peso frontalmente con velocidades *v* y –*v*, o cuando uno de ellos está parado y el otro lo embiste a *2v*? Afirman que, debido a que la diferencia de velocidades es la misma, ambos resultarían igualmente graves. Dos alumnos de física lo oyen. Las fórmulas de la energía cinética (*véase el recuadro* “Como recordatorio”) les dictan un criterio distinto: en el segundo caso la suma de energías de ambos coches dobla el valor de la suma de energías del primer caso; por ello el accidente es más grave. ¿Quién tiene razón?

La intensidad del choque en sí tiene poco que ver con las velocidades medidas con respecto a la carretera. La energía “aprovechable” para la destrucción y deformación es la energía cinética en el sistema del centro de masas, igual en ambos casos. Hasta aquí tienen razón los agentes. Podría haber una diferencia, en cambio, en los efectos secundarios del choque: en el segundo caso, el centro de masas común se sigue moviendo, amortiguado por la fricción con la carretera; los coches pueden salirse de la calzada y chocar contra árboles o casas. Si alguien colisiona por detrás a 150 km/h contra un coche que va a 140 km/h, no será más peligroso que un impacto a 10 km/h contra un coche parado, suponiendo que no ocurra en la autopista, sino sobre un gran lago de sal sin obstáculos alrededor.

Por la misma razón, en la física de partículas elementales es decisiva la energía que tengan las partículas que chocan en el sistema de su centro de masas. Un acelerador colisionador, que dispara unas partículas elementales frontalmente contra otras, logra energías de colisión mayores que un acelerador que dispara las partículas contra un objetivo en reposo. Si bien en la teoría de la relatividad la relación entre energía y momento es distinta que en la física clásica, el resultado es cualitativamente muy parecido.

Una pelota sobre la otra: Con las dos manos deje caer desde una altura de unos decímetros, sobre una mesa, dos pelotas de peso claramente distinto y con la mejor elasticidad posible; hágalo de manera que al soltarlas la ligera esté justo encima de la pesada. La ligera producirá un gran rebote; sobrepasará incluso su altura inicial (*figura 1*). Parece que recibe energía de la pesada; pero, ¿cómo?

Puede entenderse sin cálculos si se toma el sistema del centro de masas común: el primer choque tiene lugar entre la pelota inferior, la más pesada, y la Tierra, mucho más pesada. El centro de masas de estos dos cuerpos no se diferencia casi del sistema del laboratorio, en el que los observadores están fijos sobre la Tierra. Si denominamos –*v* a la velocidad al final de la caída libre, la pelota rebotará hacia arriba tras el contacto con la mesa (mejor dicho, con el planeta) y subirá con una velocidad *v*. La más ligera estará todavía de camino con velocidad –*v*. Ambas sólo entrarán en contacto estrecho cuando la de abajo pierda el suyo con el planeta. Si suponemos idealmente que la masa de la pelota pequeña es despreciable, la más pesada estará en reposo con respecto al sistema del centro de masas de ambas pelotas; la más ligera viene desde arriba con una velocidad –*2v*,

COMO RECORDATORIO

Para el momento *p* y la energía cinética ***E*** de una masa puntual *m*, que se mueve a velocidad ***v***, valen las fórmulas

$\boldsymbol{p} = m\boldsymbol{v}$

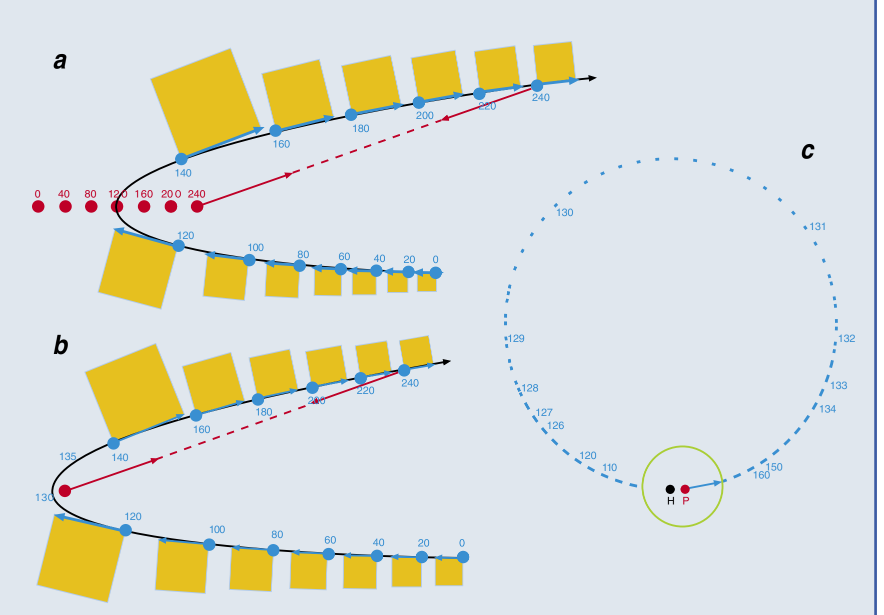
y

$E = (m/2)\ v^2$

v y ***p*** son vectores, lo que se ha de tener en cuenta a la hora de sumar momentos. Con *v* nos referimos al módulo del vector ***v***.

El impulso por aproximación

Una sonda espacial (*azul*) vuela casi frontalmente contra un planeta (*rojo*) y lo rodea en una hipérbola muy ceñida. La gráfica *a* muestra la trayectoria en el sistema de referencia heliostático (en el que el Sol está parado). La gráfica *b* ofrece la misma trayectoria en el sistema del centro de masas del planeta y la sonda espacial; en razón de la masa escasa de ésta es casi idéntico al sistema de referencia del planeta. Los cuadrados amarillos pegados a los vectores de velocidad representan la energía cinética de la nave espacial, medida en el correspondiente sistema de referencia; los números azules designan los instantes temporales en unidades arbitrarias.



La gráfica *c* muestra la velocidad de la nave espacial (puntos finales de los vectores de velocidad) en diferentes instantes del impulso por aproximación. En esta representación ("espacio de velocidades") los dos sistemas de referencia que consideramos aquí se distinguen sólo por la posición de su origen, *H* y *P* respectivamente; la trayectoria de la sonda espacial es un círculo que pasa por *P* en el caso límite de la trayectoria parabólica. El pequeño círculo verde alrededor de *P* contiene todas las velocidades correspondientes a la energía cinética de "antes" y "después" del choque en la "infinita lejanía".

visto en el sistema de ese centro de masas, y debido al contacto invierte su dirección; por tanto, luego tiene $+2v$ en ese sistema y $+3v$ en el del laboratorio. Con esto se eleva hasta nueve veces su altura inicial, ya que la energía cinética va como el cuadrado de la velocidad y la energía potencial gravitatoria es en la práctica proporcional a la altura (para esas alturas pequeñas cabe considerarlo un "campo homogéneo").

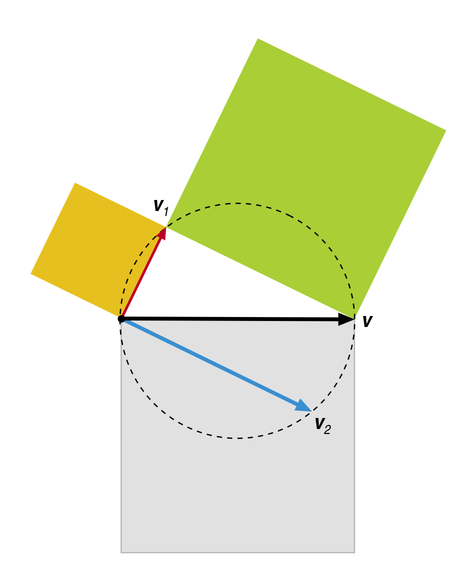
Hay que calcular algo más si la pelota ligera no es infinitamente ligera. Si usted tiene dos pelotas de buena elasticidad con una proporción de masas 1 a 3, conseguirá una transferencia completa de energía cinética medida en el sistema del laboratorio: tras el primer impacto presentan también velocidades $+v$ la pelota pesada y $-v$ la ligera, pero el centro de masas sube con una velocidad $+v/2$. Respecto al centro de masas, la pelota pesada porta entre los dos impactos una velocidad $+v/2$ y la ligera $-3v/2$. Ambas sólo cambian de signo tras el segundo choque; el cálculo en el sistema del laboratorio da cero para la pesada y $+2v$ para la ligera. Esta puede, pues, subir hasta el cuádruple de su altura inicial, mientras que la pesada, aun sin ningún tipo de fricción, no tarda en estar en el suelo. En el idealizado primer caso subiría, sin embargo, como si no existiera la pelota ligera: hasta la altura desde la que cayó.

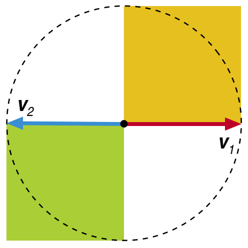
La energía cinética en el sistema del laboratorio era antes de ambos choques $2mv^2$, si denominamos a las masas m y $3m$. Tras los dos choques la pelota más ligera dispone para sí sola de toda esa energía.

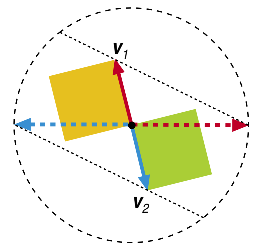
Los planetas como donantes de momento: Supongamos que queremos enviar una cámara de televisión, con todos sus aditamentos, a un planeta lejano para observarlo de cerca. Desgraciadamente, el pozo del potencial del Sol es demasiado profundo y nosotros giramos con la Tierra. Si la nave espacial se mueve heliostáticamente (es decir, con respecto a un sistema inercial en el que el Sol esté quieto) a 10 km/s, tendrá una velocidad respetable, pero no bastará. Si se lanzara el objeto a 10 km/s contra un "planeta" suficientemente elástico y acolchado, su velocidad tras el rebote portará, con respecto al astro, otra dirección y la misma magnitud. Pero si el mullido cuerpo celeste se halla en movimiento, la velocidad heliostática de la nave espacial cambiará considerablemente.

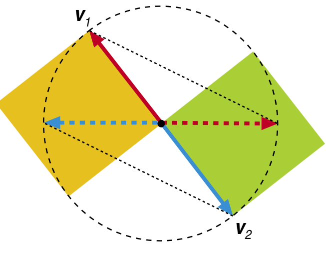
Pero los planetas no están acolchados, ni valen como reflectores de naves espaciales. Aquí nos ayuda un truco: para un cambio brusco de dirección en el patio del colegio no hay que estamparse contra una pared, sino girar alrededor de una columna o un amable compañero. Los brazos originan una atracción. En el espacio, los planetas nos proporcionan su campo de atracción gravitatorio. Si la nave espacial apunta justo a un lado del planeta, describirá una hipérbola en el sistema del centro de masas de ambos cuerpos. De esa figura sólo necesitamos las asíntotas, como aproximación del recorrido de ida y de vuelta. Si el planeta no es un mero punto, un cambio de dirección de 180 grados no será posible: el vértice de la trayectoria tendría que hundirse en el planeta. Alejándonos mucho por las asíntotas tendremos, a distancias iguales, magnitudes iguales de la velocidad de la nave espacial, medidas con

NORBERT TREITZ/SDW









2. En un choque elástico entre dos bolas de billar, representado en el espacio de las velocidades (*a la izquierda en el sistema del laboratorio, a la derecha en el del centro de masas*), la energía cinética total final es la misma que al principio, como se deduce del teorema de Pitágoras. Sólo durante el choque habrá empleada en deformar las bolas una parte de la energía.

respecto al planeta. Pero a nosotros nos interesan las velocidades heliostáticas.

Como en el caso del choque entre pelotas, se muestra aquí la fuerza decisiva del balance de energías y momentos: podemos determinar las velocidades de "antes" y "después" sin preocuparnos de los detalles. Ni siquiera debemos saber si se trata de una atracción o de una repulsión. Supongamos como velocidades heliostáticas iniciales –10 km/s para la nave espacial y +20 km/s para el planeta; así, la velocidad relativa antes y después será de 30 km/s. En el caso de un cambio de dirección completo, la nave espacial seguirá viaje a una velocidad de casi 50 km/s. Este es el principio que permitió la larga travesía de las sondas Voyager a través de todo el sistema solar. La sonda Cassini alcanzó el año pasado el sistema de anillos de Saturno tras haber ganado momento adicional en Venus (dos veces), la Tierra y Júpiter.

En general, el objeto más ligero puede ganar casi el doble de la velocidad que el más pesado; esto vale para cualquier sistema inercial. Por idéntica razón, la velocidad de una pelota de tenis inicialmente en reposo duplicará la velocidad de la raqueta.

Billar pitagórico: Las bolas de billar pesan lo mismo. Por eso, la adición de sus momentos se convierte en la adición de sus velocidades. La representación en el espacio de las velocidades, que se explica en el recuadro "El impulso por aproximación", nos ofrece aquí de forma sorprendentemente rápida una clara imagen.

Una bola de billar alcanza con una velocidad $\boldsymbol{v}$ a una bola en reposo sobre la mesa. En el espacio de las velocidades las bolas se hallan antes del choque en el sistema del centro de masas a una distancia $v/2$ del origen. De nuevo será así tras un choque elástico (pero no durante el mismo). Tras el choque, las velocidades de salida —en esa representación— estarán dispuestas, sobre radios opuestos, en el mismo círculo de radio $v/2$ centrado en el origen. (En un movimiento en tres dimensiones, el círculo debe sustituirse por una esfera.) Denominemos $\boldsymbol{v}_1$ y $\boldsymbol{v}_2$ a las nuevas velocidades; así los puntos 0, $\boldsymbol{v}$, $\boldsymbol{v}_1$ y $\boldsymbol{v}_2$ forman las esquinas de un rectángulo (*figura 2*).

Pasar del centro de masas al sistema del laboratorio —esto es, el sistema en el que la mesa de billar está en reposo— equivale a un desplazamiento del origen en el espacio de las velocidades. El vector $\boldsymbol{v}$, la velocidad inicial de la bola que impacta, es ahora la diagonal del rectángulo o, si se divide por la mitad el rectángulo a lo largo de esta diagonal, la hipotenusa de un triángulo rectángulo. Sus catetos son las velocidades de "después".

De esta manera tan simple se ve no sólo que las bolas siguen trayectorias perpendiculares entre sí tras el choque, sino que también se vuelve a alcanzar la energía cinética total. (Estrictamente hablando, no se conserva, porque durante el choque se almacena energía elástica.) Según el principio de conservación del momento, $\boldsymbol{v}$ debe ser la suma de $\boldsymbol{v}_1$ y $\boldsymbol{v}_2$, esto es, los tres vectores deben formar un triángulo, que por el teorema de Pitágoras será rectángulo, pues el balance de energías da (debido a que las masas son iguales) $v^2 = v_1^2 + v_2^2$.

Durante el choque, las puntas de las flechas $\boldsymbol{v}_1$ y $\boldsymbol{v}_2$ se mueven sobre líneas rectas desde su posición inicial hasta la final. Mientras, el triángulo no es rectángulo y la suma de los cuadrados de las velocidades no llegan transitoriamente a v^2, pues una parte de la energía se almacena como energía de deformación. La dirección del impulso transmitido la hemos considerado hasta aquí casual. Coincide simplemente con la dirección resultante de unir los dos centros de las bolas, que suponemos bien redondas, durante el contacto. Una bola que se aleje del choque casi en dirección hacia delante recibe casi toda la energía; una que se aleje casi perpendicularmente, apenas ninguna. Qué suerte corre cada bola dependerá de si se trata de un choque central o de un mero roce.

Probablemente, Pitágoras de Samos no jugaba al billar. Es una pena, porque habría encontrado una bonita aplicación al teorema que lleva su nombre: la redistribución de energía cinética en el choque de las bolas de billar.

SPEKTRUM DER WISSENSCHAFT

JUEGOS MATEMÁTICOS

Juan M.R. Parrondo

La forma de un iceberg

Todos sabemos que la mayor parte de la masa de un iceberg se encuentra sumergida. La "punta del iceberg", es decir, su parte visible, es algo menos de un 10 % del volumen total.

Hace unos meses, paseando por un río parcialmente helado, comencé a lanzar bloques de hielo al agua. Pronto me di cuenta de que ninguno de ellos se parecía a la imagen de la figura 1. Intenté esculpir el hielo para que adquiriera la forma del iceberg prototípico pero, al introducirlo en el agua, el bloque se orientaba siempre con su lado más ancho hacia la superficie. En otras palabras, un trozo de hielo como el de la figura 1 giraría unos 90 grados hasta que el lado más largo quedara horizontal. Según mis "experimentos" la imagen clásica del iceberg es imposible. ¿Es esto cierto?

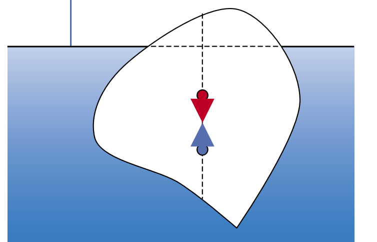


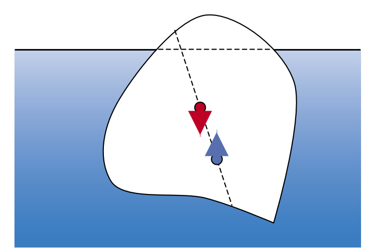
1. Creación artística de un iceberg.

Todos nos imaginamos a Arquímedes envuelto en una toalla y gritando Eureka por las calles de Siracusa. Sabemos que un cuerpo sumergido en agua experimenta una fuerza o empuje hacia arriba igual al peso del agua que desaloja. Si un cuerpo de volumen V tiene una densidad d inferior a la del agua (supondremos que la densidad del agua es igual a 1 gr/cm^3 y mediremos la densidad d del cuerpo en estas unidades), el cuerpo flotará. Cuando el cuerpo llega a la superficie, se alcanza una situación de equilibrio en la que parte de su volumen sale al exterior y queda sumergido un volumen $V_{sum} < V$. En la situación de equilibrio, el peso del cuerpo Vd tiene que ser igual al empuje V_{sum} y, por tanto, la fracción de volumen sumergido V_{sum}/V coincide con la densidad d del cuerpo expresada en gramos por centímetro cúbico. En el caso del hielo, la densidad es aproximadamente 0,9. Por eso la "punta del iceberg" es sólo un 10 % de su volumen total.

La parte difícil del problema es encontrar la orientación que adquiere el cuerpo flotante, una cuestión a la que han dedicado esfuerzos muchos físicos, de Huygens a Sommerfeld. Para resolverla, debemos recordar que el peso de un cuerpo se puede considerar una fuerza aplicada en su centro de masas o centro de gravedad. El peso es en realidad una fuerza que la tierra ejerce sobre todas y cada una de las moléculas del cuerpo, pero se puede demostrar que, a todos los efectos, equivale a una fuerza aplicada en dicho punto. Del mismo modo, el empuje se puede considerar una fuerza aplicada *en el centro de gravedad de la parte sumergida*.

La condición de equilibrio para nuestro cuerpo flotante es sencilla: el centro de gravedad del cuerpo y el centro de gravedad de la parte sumergida tienen que estar en la misma vertical. Si no fuera así, el empuje y el peso harían girar al cuerpo. Esta situación puede verse en la figura 2. En azul se muestra el centro de gravedad de la parte sumergida y el empuje, y en rojo el centro de gravedad del cuerpo y su peso. Si los dos centros de gravedad no están en la misma vertical (*dibujo de la derecha*), el peso y el empuje harán girar al cuerpo.





2. La "paradoja" del cuerpo flotante: si giramos el cuerpo hacia la izquierda, aparece un par de fuerzas que hace que el cuerpo gire también hacia la izquierda. ¿Es entonces inestable cualquier orientación del cuerpo?

Sin embargo, en el problema de la orientación del cuerpo hay algo que da una impresión paradójica. Supongamos que giramos ligeramente hacia la izquierda un cuerpo en equilibrio, como en la figura 2. El eje que forman los dos centros de gravedad, el del cuerpo y el de la parte sumergida, habrá girado hacia la izquierda, con lo que el centro de gravedad de la parte sumergida (*punto azul*) quedará a la derecha del centro de gravedad del cuerpo (*punto rojo*). Como se ve en la figura, el peso y el empuje forman ahora un par de fuerzas que hace girar al cuerpo hacia la izquierda, alejándolo de su posición de equilibrio. El argumento puede repetirse para un giro hacia la derecha. Por lo tanto, cualquier pequeña desviación de la orientación de equilibrio crea un par de fuerzas que lo aleja más aún de dicha orientación. Sin embargo, basta observar un cubo de hielo flotando en un vaso de agua para convencerse de que esto no es cierto. Golpee el hielo para que gire un poco y verá que enseguida recupera su situación original. Un cuerpo flotante tiene siempre al menos una orientación de equilibrio estable. ¿Cuál es entonces el error de nuestro argumento? Es bastante simple: hemos supuesto que el centro de gravedad de la parte sumergida gira igual que el cuerpo, algo que no tiene por qué ser cierto, ya que la propia geometría de la parte sumergida cambia. En una orientación de equilibrio estable, el centro de gravedad de la parte sumergida se mueve hacia la izquierda del centro de gravedad del cuerpo cuando éste gira hacia

FOTOMONTAJE DE UWE KILS, http://en.wikipedia.org/wiki/Iceberg (*arriba*)

la izquierda; se crea así un par que hace que el cuerpo recupere su orientación de equilibrio. Veamos un caso particular.

Analicemos cómo se puede orientar un rectángulo de lados *a* y *b* y densidad *d*. En la figura 3 se ofrece una posible orientación de equilibrio, en donde los centros de gravedad del cuerpo y de la parte sumergida están alineados verticalmente. Esta no es la única orientación de equilibrio: girando el rectángulo 90 grados, encontraríamos otra orientación posible, cuyo análisis es idéntico, intercambiando los lados *a* y *b*. El centro de gravedad del cuerpo (*punto rojo*) está situado en la mitad del rectángulo, es decir, a una altura *b*/2 con respecto a su base. La parte sumergida del rectángulo es de nuevo un rectángulo de altura *bd*, puesto que, así, la fracción de volumen sumergido es igual a la densidad del cuerpo *d*. Por lo tanto, el centro de gravedad de la parte sumergida (*punto azul*) está a una altura *bd*/2. La distancia entre los dos centros de gravedad es entonces $b(1-d)/2$, tal y como se indica en la figura 3.

¿Cómo cambian los centros de gravedad si giramos ligeramente el cuerpo hacia la derecha? El giro está representado en la figura 4 desde el punto de vista del cuerpo. Es decir, en la figura he girado la superficie del agua hacia la izquierda. Después del giro, el triángulo rayado, antes sumergido, deja de estarlo; otro triángulo igual, el sombreado en gris, se sumerge, de modo que la fracción de volumen sumergida no cambia, pero sí su forma. En consecuencia, el centro de gravedad de la parte sumergida se desplaza hacia la derecha una cantidad *l*, como se indica en la ampliación del dibujo. El nuevo centro de gravedad se puede calcular como el promedio, convenientemente ponderado, de tres partes: la parte que estaba sumergida al principio, el triángulo rayado (con peso negativo) y el triángulo gris. Con ello se obtiene:

$$l = 2 \times \frac{a/3 \times A_{\Delta}}{abd} = \frac{as}{6bd},$$

donde $A_{\Delta} = as/4$ es el área del triángulo rayado; hemos utilizado el hecho de que los centros de gravedad de los triángulos (*puntos verdes en la figura 4*) están a una distancia *a*/3 del eje vertical. Para que la orientación inicial sea de equilibrio estable, se requiere que el par de fuerzas que aparece después del giro sea hacia la izquierda, de modo que tienda a restituir la orientación inicial. Eso ocurrirá si el nuevo centro de gravedad de la parte sumergida queda a la derecha del centro de gravedad del cuerpo, es decir, si $l > r$, ya que la línea naranja de la figura 4 constituye la nueva vertical. La condición de estabilidad es entonces:

$$\frac{as}{6bd} > r \Rightarrow \frac{a}{6bd} > \frac{r}{s}.$$

Finalmente, el cociente *r/s* puede expresarse en términos de los datos del problema si se tiene en cuenta que el triángulo rayado y el verde son equivalentes. Así que:

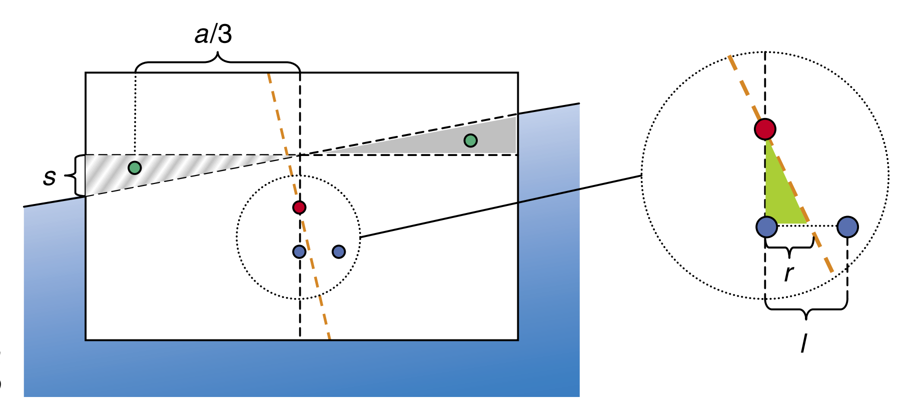
$$\frac{r}{s} = \frac{b(1-d)/2}{a/2} = \frac{b(1-d)}{a}$$

y la condición de equilibrio es:

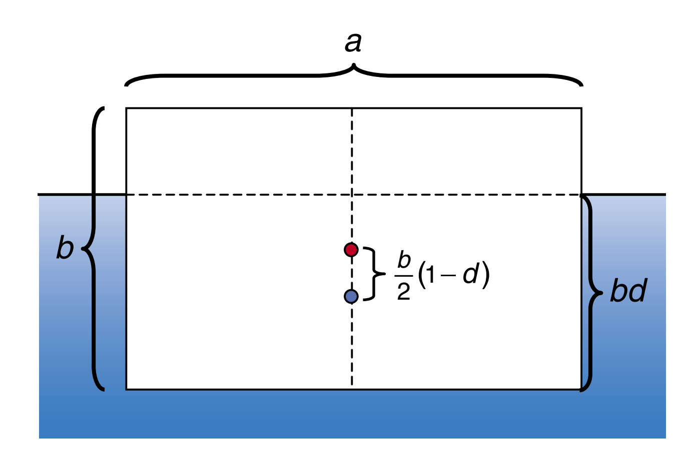
$$\left(\frac{a}{b}\right)^2 > 6d(1-d).$$

Para el hielo ($d = 0,9$) esta condición es $a/b > 0,73$. Es decir: cualquier orientación con $a > b$ es estable. Pero un rectángulo en el que la relación entre el lado menor y el mayor sea superior a 0,73, se puede orientar de forma estable tanto con su lado mayor horizontal como vertical. Dependiendo del valor de *d*, pueden aparecer otras orientaciones estables, e incluso asimétricas, con sólo una o tres de las esquinas del triángulo sumergidas.

Sin embargo, la anchura del bloque de hielo de la figura 1 es aproximadamente la mitad de su altura. Por lo tanto, la orientación del fotomontaje no es estable, al menos si el iceberg es homogéneo. Debo indicar que el problema de la orientación de un iceberg reviste mayor complejidad que la abordada aqui. Depende incluso de su historia geológica. Por ejemplo, se sabe que la parte sumergida se funde más rápido que la visible; por ello, el iceberg puede volcar de modo repentino. No obstante, creo que nuestro análisis ofrece razones suficientes para dudar de la imagen que la mayoría nos hacemos de estos grandes bloques de hielo.



4. Cómo cambia el centro de gravedad de la parte sumergida cuando se gira ligeramente el cuerpo hacia la izquierda.



3. Una posible situación de equilibrio para un rectángulo flotante.

Mark Fischetti

SIN LLAVE

Abrete, sésamo

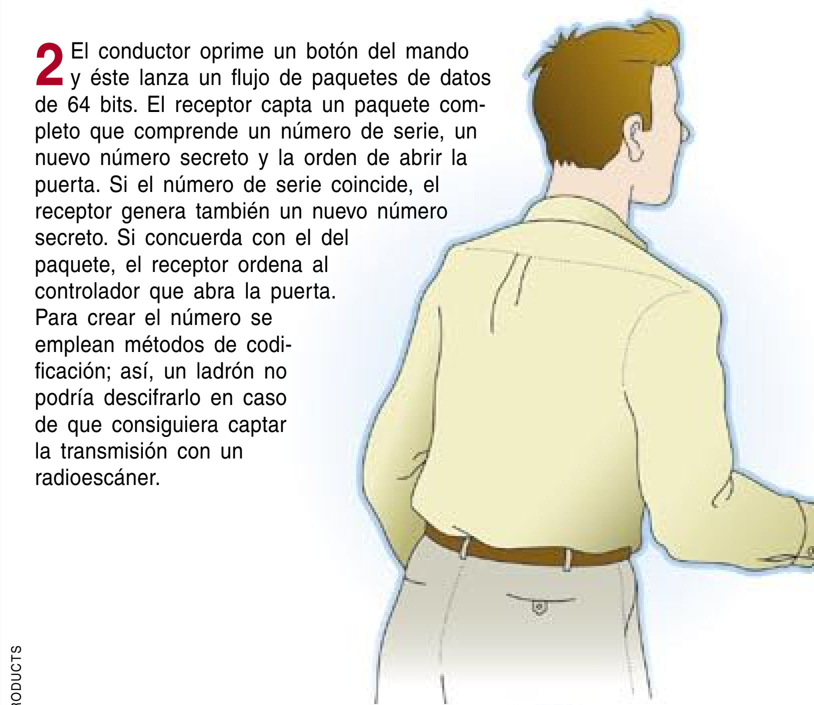
Sale usted de un hipermercado muy concurrido, apunta su mando a distancia hacia una parte del atestado estacionamiento, aprieta un botón y la puerta de su automóvil se abre. ¿Cómo sabe el vehículo que es usted? ¿Y por qué no se abren también otros automóviles?

Casi todos los vehículos nuevos ofrecen el "acceso remoto sin llave", como dicen en la industria del automóvil. Los fabricantes programan el microprocesador instalado en cada vehículo y el mando a distancia con una fórmula algorítmica que genera un código cifrado secreto que varía cada vez que se oprime el botón (*véase ilustración*). Las fórmulas son únicas, de modo que cada vehículo sólo puede abrirse con el mando que le corresponde. Las puertas de los garajes modernos funcionan con un mecanismo similar. "El sistema lleva diez años en el mercado y se ha mostrado muy seguro", dice Francis Duvenhage, director comercial de Microchip Technology, empresa de Chandler, Arizona, con cuyos procesadores codificadores funcionan la mayoría de los mandos a distancia electrónicos de los automóviles norteamericanos. "Ningún ataque al algoritmo ha tenido éxito".

El éxito ha propiciado unas comodidades y una protección aún mayores. En la "apertura pasiva sin llave", mientras el conductor tira de la manija un transceptor (un transmisor-receptor) instalado en el vehículo contrasta el código del mando o la tarjeta inteligente que el conductor lleva en el bolsillo o en el bolso; resulta muy cómodo para quien vaya cargado de bolsas. Los sistemas de "inmovilización" impiden que el motor arranque tras introducir la llave de contacto a menos que el microcircuito incorporado en la cabeza de la llave proporcione el código correcto al microcircuito existente en la carcasa del contacto. En los países europeos que obligan a que los automóviles nuevos lleven inmovilizadores, el robo de los últimos modelos ha disminuido en más del 50 por ciento.

Las señales de comunicación en los inmovilizadores y sistemas de apertura pasiva utilizan una frecuencia bastante baja, 125 kHz. Los sistemas sin llave remotos recurren a frecuencias mucho más altas, 315 Mhz en EE.UU. y Japón, en parte porque así se abarcan áreas de aparcar enteras. Los fabricantes europeos han estado empleando 433 Mhz, observa Alec Makdessian, director comercial de Maxim Integrated Products, fabricante de circuitos integrados de Sunnyvale (California), "pero se están pasando a los 868 Mhz porque las bandas más bajas están congestionadas." Saturación a la que contribuyen del orden de dos millones de nuevos artefactos cada año en todo el mundo.

1 El fabricante del vehículo programa un receptor y un mando a distancia con un algoritmo informático; cada vez que se activan, el algoritmo genera un número de 32 bits. Ese número se crea conforme a un patrón prefijado que se determina mediante una fórmula compartida sólo por el vehículo y el mando, de modo que el vehículo sólo pueda abrirse con ese mando y no con otro.



2 El conductor oprime un botón del mando y éste lanza un flujo de paquetes de datos de 64 bits. El receptor capta un paquete completo que comprende un número de serie, un nuevo número secreto y la orden de abrir la puerta. Si el número de serie coincide, el receptor genera también un nuevo número secreto. Si concuerda con el del paquete, el receptor ordena al controlador que abra la puerta. Para crear el número se emplean métodos de codificación; así, un ladrón no podría descifrarlo en caso de que consiguiera captar la transmisión con un radioescáner.



SAMUEL VELASCO; FUENTE: COILCRAFT, INC.; MAXIM INTEGRATED PRODUCTS

¿SABIA USTED QUE...?

➤**NEUMATICOS:** Unos neumáticos con la presión baja vuelven peligrosa la conducción y pueden sufrir un reventón. Unas leyes federales recientes ordenan en Estados Unidos que, en el futuro, los nuevos vehículos estén equipados con sistemas de monitorización de la presión de los neumáticos. La opción preferida, que ya se ofrece en algunos modelos, es un sensor alojado dentro del neumático, detrás del vástago de la válvula. Si la presión baja en un 25 por ciento del valor de inflado recomendado, el radiotransmisor del sensor, alimentado por la batería, advierte al ordenador del vehículo, que enciende un icono de aviso en el salpicadero.

➤**GARAJES:** Los primeros sistemas de apertura automática de garajes aparecieron en la década de 1950. Se limitaban a enviar una orden de “abrir” o “cerrar” en una frecuencia única. Cuando poco a poco fueron proliferando, cualquiera podía abrir las puertas de los demás. Hacia los años setenta, el mando y el controlador de la puerta tenían cada uno un conmutador de circuito integrado con ocho pequeñas patillas, cuya concordancia con uno de los 256 códigos posibles se ajustaba manualmente, un sistema mejor pero aún no muy seguro. Los sistemas modernos se basan en los mismos algoritmos electrónicos que los mandos de acceso remoto: un código de 32 bits ofrece más de cuatro mil millones de combinaciones posibles.

➤**Y LAS CASAS, TAMBIEN:** En muchos hogares hay ahora cerraduras que se abren con mandos a distancia. Suelen exhibir una luz roja cuando la puerta está bloqueada. Se acabó el tener que rebuscar las llaves cuando se tienen las manos ocupadas y el tantear a oscuras en busca del ojo de la cerradura.



LIBROS

Una cuestión recurrente

Vivo e inerte

A CULTURAL HISTORY OF MEDICAL VITALISM IN ENLIGHTENMENT MONTPELLIER, por Elizabeth A. Williams. Ashgate; Aldershot, 2003. SCIENCE AND POLITY IN FRANCE. THE REVOLUTIONARY AND NAPOLEONIC YEARS, por Charles Coulston Gillispie. Princeton University Press; Princeton, 2004.

THINKING ABOUT BIOLOGY, por Stephen Webster. Cambridge University Press; Cambridge, 2003. THE PHILOSOPHY OF BIOLOGY. AN EPISODIC HISTORY, por Marjorie Grene y David Depew. Cambridge University Press; Cambridge, 2004. WHAT MAKES BIOLOGY UNIQUE. CONSIDERATIONS ON THE AUTONOMY OF A SCIENTIFIC DISCIPLINE, por Ernst Mayr. Cambridge University Press; Cambridge, 2004.

¿Difiere de lo inerte lo vivo? Si difiere, ¿en qué? Puesto que la vida constituye el objeto de estudio de la biología, habrá de divergir de las ciencias que estudian la materia inerte y sus cambios. Hablamos de su autonomía, de su reducción *cum mica salis* a la química y a la física. De esos supuestos metabiológicos se ocupa centralmente la filosofía de la biología (*Thinking About Biology, The Philosophy of Biology* y *What Makes Biology Unique*), asunto al que le dedicó Ernst Mayr con mayor voluntad que dominio los últimos años de su vida centenaria.

Si la vida difiere de lo inerte, significa que los fenómenos biológicos no pueden explicarse por las leyes ordinarias de la física y la química. Con todo, la separación, que parece inmediata, distó desde antiguo de poseer lindes fijos. No sólo se trataba de una nebulosa correspondencia entre el macrocosmos y el microcosmos en una Tierra considerada en sí misma un organismo, sino que la propia cadena de los seres y la teoría de los grados del ser consideraron un tránsito fluido el paso del barro al animal a través del origen espontáneo de formas complejas de vida.

En rigor lo que se entiende por tesis vitalista, o vitalismo, surgió como un movimiento de reacción contra el mecanicismo cartesiano (*A Cultural History of Medical Vitalism in Enlightenment Montpellier* y *Science and Polity in France. The Revolutionary and Napoleonic Years*). El pensamiento dualista de Descartes, el del organismo considerado una máquina, se prolonga en la medicina iatromecánica, que buscaba leyes universalmente válidas para todos los fenómenos. Una pretensión que les parecía imposible a los médicos vitalistas formados en Montpellier, que postulaban la separación absoluta y tajante entre lo vivo y lo inerte.

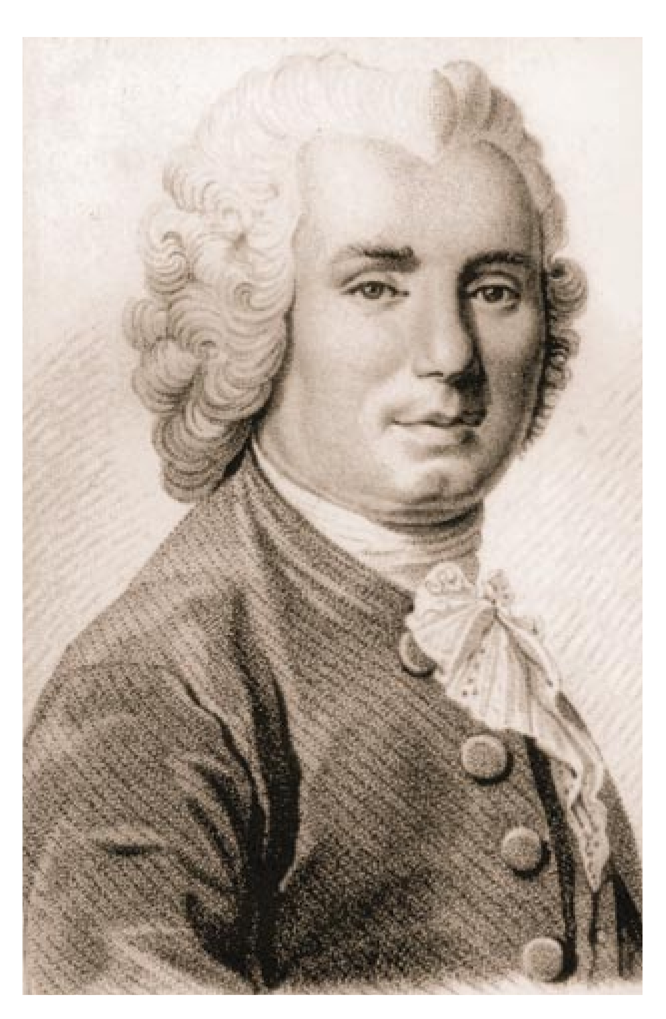
Los vitalistas atribuyeron la vida a la acción de una fuerza, principio o potencia cuyo origen y estatuto ontológico era incognoscible. De acuerdo con uno de sus precursores, J. B. Van Helmont, ese principio era un arqueo exógeno. Los teóricos de Montpellier, en particular, Théophile de Bordeu y Paul-Joseph Barthez, abogaron por una fuerza endógena. La vida consistía en las actividades interrelacionadas de la "economía del cuerpo" y se encontraba comprometida en una batalla constante contra los procesos de desequilibrio, ruptura y desintegración, que constituían la enfermedad. Hablaron de interacciones incesantes entre la disposición interna y el medio externo que acaecían en distintos tipos humanos formados por la edad, sexo, temperamento, región y otras influencias poderosas sobre la economía vital.

El antagonismo entre mecanicismo y vitalismo puede esquematizarse en una tabla de opuestos que, no obstante su obligada simplificación, facilita la comprensión: leyes universales de la naturaleza muerta frente a leyes particulares de la naturaleza viva; ciencia determinista frente a ciencia integrada en el mundo indeterminado de lo vivo; ciencia fundada en la matemática frente a ciencia antimatemática; experimento y observación frente a observación sólo; unidad de las ciencias frente a autonomía de la medicina; explicación antiteleológica frente a explicación teleológica.

La escuela de Montpellier no puede entenderse sin conocer el impulso recibido del nosólogo François Boissier de Sauvages, espíritu paradójico que supo mantener la tradición médica protestante de Montpellier, reforzada a través de su intensa actividad epistolar con Uppsala, Londres, Halle, Ginebra y Lyon, con una observancia católica que dio, en su propio seno familiar, vástagos a la Iglesia. Nombrado sustituto de Eustache Marcot en 1734, se hizo cargo del jardín real en 1740. La facultad contaba con aulas, anfiteatro, jardín y desde 1720 laboratorio de química. Había ocho cátedras, profesores ordinarios y demostradores. El jardín botánico había sido establecido en 1593. Entre 1740 y 1748, el intendente del jardín fue Aimé-François Chicoyneau, quien se había formado en París con Sébastien Vaillant, descubridor de la sexualidad vegetal y discípulo de Joseph de Tournefort.

Montpellier participaba también de la inquietud intelectual de la mesocracia provinciana con instituciones que recibieron la sanción real. La Sociedad Real de Ciencias de Montpellier se había fundado en 1706, promovida por un grupo de notables locales que contaron con la ayuda de Gian Domenico Cassini, quien había bajado al sur en 1700-1701 dentro de su proyecto de extender el meridiano de París. Aunque se habían establecido ya las de Nimes, Angers, Caen y Toulouse, la de Montpellier fue la primera en dedicarse al cultivo de la ciencia, y la primera en recibir un estatuto de igualdad con la de París.

Desde su incorporación al claustro docente, Sauvages rompió con la doctrina iatromecánica y iatroquí-



Théophile de Bordeu (Izeste, 1722-París, 1766).

mica, para promover tesis "animistas" que engarzaban con las enseñanzas del controvertido profesor de Halle Georg-Ernst Stahl. No es fácil clasificar la medicina de Sauvages, una mezcla de elementos mecanicistas, animistas y vitalistas en tensiones sin resolver. Mayor reconocimiento ha merecido su empeño sistematizador de las enfermedades, a imagen de la taxonomía botánica, materia en la que adquirió notoriedad.

En su enfrentamiento al mecanicismo, el Sauvages de *Nouvelles classes des maladies* dependía también de los médicos newtonianos ingleses, quienes, gráficamente, se declaraban contrarios a la pretensión cartesiana de explicarlo todo apelando a los torbellinos. Parecía manifiesto que había fuerzas o causas no analizables, indemostrables. Bryan Robinson, Stephen Hales y Browne Langrish, entre otros, daban por supuesto que el movimiento del corazón y otras funciones cruciales estaban controladas por el alma, que actuaba a través de fuerzas operativas del organismo. El enfoque newtoniano se combinó pronto con las tesis de Stahl. En esa línea gozó de particular éxito el texto publicado en 1751 por Robert Whytt, *Essay on the Vital and Involuntary Motions of Animals*, donde postulaba un alma que fuera coextensa con el cuerpo y que en cada órgano ejerciera las tareas vitales asignadas a dicha estructura.

En línea con ese pensamiento, Sauvages rechazaba que el cuerpo funcionase como un mecanismo ciego. Apoyándose en las fuerzas de "gravedad" y "adhesión" explicaba por qué las medicinas obedecían reglas y actuaban de una manera predictible. Articula sus ideas en la *Nosologie*, donde polemiza con las "sectas" que habían sembrado la confusión en medicina, el mecanicismo en particular; le recriminaba incluso que "sus principios distaban mucho de ser verdaderos principios de la mecánica". Con el término *nosología* designaba la ordenación jerárquica de las enfermedades, de acuerdo con un triple criterio: establecer una división exacta y puramente histórica de los géneros y especies de las enfermedades; distinguir teoría filosófica o hipótesis de la historia, y definir el carácter de las enfermedades mediante síntomas constantes. Las enfermedades constituían entidades específicas, con propiedades discernibles y estables.

Aunque adopta el concepto stahliano de "alma", acostumbra asociarlo con la noción de fuerza. Su significado, sin embargo, resulta bastante inestable. A menudo lo emplea en el sentido que recibe en la dinámica newtoniana; con otras palabras, los movimientos del cuerpo podían atribuirse a la gravedad, elasticidad y tenacidad, fuerzas perceptibles en los cuerpos vivos e inertes. También algunos fenómenos patológicos podían explicarse mediante fuerzas mecánicas universales. Las heridas, por ejemplo, constituían el efecto de la fuerza que separa partes continuas. Distinguía entre fuerzas mecánicas, mensurables y claramente definidas, y fuerzas como potencias motoras de órganos específicos del cuerpo y en general de la máquina viva. Unía, pues, el organismo humano a las leyes generales del universo. Pero dio un paso más: de la función de las fuerzas a la función de las facultades. Estas eran propiedades del alma, peculiares de criaturas vivas específicas. Todos los seres vivos tenían "almas" dotadas de ciertas facultades. Las facultades del alma eran tres: conocer, desear y tender a.

Uno de los integrantes de primera hora del círculo vitalista de Montpellier fue Théophile de Bordeu. Cuando se inscribió en la facultad de medicina, en 1739, encontró un plantel docente muy decoroso. François Chicoyneau enseñaba anatomía y botánica; Henri Haguenot, cirugía y farmacia; química, Antone Fizes. Con España tendría que ver Antoine Magnol, quien terminó y publicó la obra de su padre *Novus character plantarum*. Sin embargo Fizes es, con Sauvages, el que parece haber ejercido mayor influencia. Tras graduarse y publicar su primer libro, un tratado sobre aguas medicinales de Béarn, Bordeu marchó a París en 1746. De acuerdo con la tesis vitalista defendía en ese texto que la salud y la enfermedad reflejaban el condicionamiento a largo plazo del cuerpo a través de los hábitos sostenidos de la dieta, el régimen y el ejercicio. Bordeu consideró siempre la enfermedad una condición crónica, no un episodio agudo. Las enfermedades ponían de manifiesto la debilidad general de la economía del cuerpo y su tratamiento requería una terapia a largo plazo. Por consiguiente el uso de las aguas minerales constituía una forma de vida. Había entonces en París un nutrido e influyente grupo de médicos procedentes de Montpellier,



Macho adulto de ave del paraíso.

entre ellos, Jean Astruc, François de la Peyronie y el propio Chicoyneau. Pero lo que un provinciano buscaba en París era introducirse en los salones de la *Encyclopédie*.

Entre 1747 y 1749, Bordeu compuso sus *Recherches anatomiques sur la position des glandes et sur leur action*, un alegato contra el mecanicismo y defensa de las tesis vitalistas.

De entrada, proponía que el conocimiento de las funciones de las glándulas debía basarse en la comprensión pormenorizada de su estructura anatómica. Para describir la función glandular, había que empezar por arrumbar la explicación mecanicista; según ésta, las glándulas operaban en respuesta a la compresión de la substancia glandular ejercida por el tejido muscular circundante. Bordeu pasó revista a la estructura y posición de las glándulas parótida, lacrimal, salival, bucal, aritenoide, pituitaria, tiroidea y timo, así como a los órganos blandos considerados de función glandular; entre ellos, el páncreas, riñón, testículos y cerebro. Tras varios experimentos con cadáveres, aporta una explicación experimental, alejada de la mecanicista, a la que consideraba "puro producto de la imaginación". En su opinión, la acción glandular se producía en respuesta a "una especie de sensación", un estado nervioso de *eretismo*, análogo al de la erección del pene en el acto procreador. La erección de la glándula conducía a una suerte de convulsión de sensaciones que terminaba con la efusión de semen, leche, lágrimas, saliva y demás humores. Para ello se requería que se desplegaran cuatro condiciones: una disposición favorable de los vasos sanguíneos a partir de los cuales derivaban los humores; una disposición idónea de los canales excretores; algún tipo de convulsión, irritación o fricción; y, por último, el espasmo que pone en movimiento a los humores.

Se ignora si Diderot había leído la obra de Bordeu cuando se decidió a publicar el artículo "Crise", la única aportación de éste último a la *Encyclopédie.* Apareció en el volumen 4, de 1753. Ese texto constituye, junto con el artículo de Gabriel-François Venel "Chymie", una suerte de manifesto vitalista. Bordeu aborda la validez de los días críticos, en los que, de acuerdo con la doctrina hipocrático-galénica, la enfermedad alcanzaba su crisis y empezaba su marcha hacia la resolución. Bordeu convirtió esta cuestión, de apariencia anodina, en un arma para atacar la tradición médica y para meditar sobre el carácter fundamental de la enfermedad. Denunciaba nuestra ignorancia de los procesos de la naturaleza, que sólo se disiparía tras una atenta auscultación de la misma en múltiples observaciones por médicos expertos.

De Montpellier procedía también Venel, y a su patria chica volvería a finales de los años cincuenta tras un decenio de actividad en París. La entrada "Chymie" se estructuraba reconociendo la superioridad intrínseca de la química sobre la física a la hora de explicar los fenómenos terrestres como opuestos a los fenómenos celestes. Al establecer esa distinción, Venel desarrollaba temas que iban a ser constantes en el discurso vitalista: distinción entre lo vivo y lo no vivo, entre lo orgánico y lo inerte; el papel desempeñado en los fenómenos químicos por fuerzas intrínsecas, aunque indefinibles; el sinsentido de intentar aprehender los fenómenos químicos en imágenes de figura, forma o tamaño, o buscar su causa en burdas acciones mecánicas como agitar, macerar o presionar. Lo mismo que Bordeu en "Crise", Venel dedicó una parte extensa de su "Chymie" a una exposición histórica de la química, donde ataca a los mecanicistas y, sobre todo, a Robert Boyle, a quien Venel acusaba de ignorancia de los fenómenos químicos. Despreció a los newtonianos John Keill, John Freind,

Stephen Hales y al propio Newton. Venel abogaba por una química verdadera, lo mismo que Bordeu apelaba a una medicina genuina. Ambos también denunciaban que los mecanicistas minusvaloraran el poder de la naturaleza, de sus fuerzas, indomeñables e incognoscibles.

Por encima de todo, Venel desdeñaba la matematización excesiva. Admitía, por supuesto, la utilidad de la matemática en física —la fuerza de atracción que operaba entre dos objetos físicos podía medirse con exactitud—, pero negaba que fuera siquiera posible medir la fuerza miscible que mantenía juntas las partículas de una reacción química. Ni siquiera tal unión podía explicarse mediante acciones mecánicas groseras, que sólo podían indicar la energía liberada por agentes genuinamente químicos.

Con Bordeu y Venel, forma Paul Joseph Barthez el famoso triunvirato vitalista de Montpellier, aunque hubo varios más. Pero si los dos primeros llegaron a la Enciclopedia a través de Diderot, el tercero fue introducido por d'Alembert y, sobre todo, por Camille Falconet, médico de la corte. Barthez llegaría a ser coeditor del *Journal des savants*. Erudito y lingüista, sus colaboraciones en la *Encyclopédie* no fueron sólo sobre anatomía (músculos flexores y extensores, estructura del rostro y temas similares), sino también sobre cuestiones clásicas e historia de la filosofía.

Los artículos "follicule" y el larguísimo "force des animaux" reflejan su adscripción vitalista. Al abordar la fuerza animal, el objetivo principal de Barthez era desacreditar las propuestas de los iatromecanicistas, en particular la idea defendida por George Cheyne, Freind y Wainwright según la cual la fuerza animal existía en proporción a la cantidad de la masa de sangre. De una manera más general, Barthez proponía que la fuerza animal podía concebirse como equivalente a la fuerza de los cuerpos inertes.

Pero ninguna obra resultó tan decisiva para culminar la articulación del pensamiento vitalista que los *Nouveaux éléments de la science de l'homme*. Para Barthez resultaba esencial la unidad del principio vital. E insistía en que el principio vital debía concebirse como un todo que pudiera modificarse por influencias diversas y en direcciones y sitios orgánicos distintos. El principio vital guiaba todas las acciones y reacciones: sentir, mover, respirar, digerir o excretar.

Los vitalistas de Montpellier ofrecían una comprensión holista de la relación físico-moral. Su medicina no se basaba en las matemáticas y en la unidad de la ciencia, sino en la observación del paciente individual y en las actividades armoniosas de la "economía corporal". Creían que la enfermedad era resultado de la falta de armonía en su "economía del cuerpo", desequilibrio que sólo podía remediarse en el sujeto, uno a uno y de acuerdo con las limitaciones "naturales" del propio paciente. Tales limitaciones quedaban establecidas por una miríada de factores: sexo, clase, edad, temperamento, región y raza; esos factores negaban el uso de un tratamiento universal para una patología particular. Por eso buscaron, los que pudieron, regresar al punto de origen. En 1776 el enfermo y agotado Bordeu, cansado del frío y de las intrigas de París, se encaminó hacia el sur para recuperar su salud y su felicidad. Henri Fouquet declaraba que esta tierra que produce y nos nutre, sana nuestras enfermedades. A la postre, la medicina de Montpellier fue eclipsada por la de París, un desarrollo asociado a la dinámica de la Ilustración, centralizadora en su cúspide napoleónica (*Science and Polity in France*).

Desde los años setenta del siglo XVIII hasta avanzados los veinte de la centuria siguiente, la ciencia francesa dominó en Europa en un grado tal, que no volvería a repetirse con ninguna otra nación antes de la Segunda Guerra Mundial. En su clásico *Science and Polity in France: The End of Old Regime*, Charles Gillispie analizaba los factores culturales, políticos y técnicos que sostuvieron la creación científica en el alba de la Revolución. En el nuevo volumen, prosigue con pareja musculatura el devenir de la interrelación de la ciencia y la política en el contexto revolucionario y napoleónico. Si en



Ernst Mayr (1904-2005).

política se trata de la conversión de los súbditos de una monarquía en ciudadanos de una república, para la comunidad científica supone la adquisición de un estatuto profesional públicamente reconocido. ¿Qué tiene ello que ver con el vitalismo?

El nuevo rumbo de la fisiología lo marca la obra de Xavier Bichat, cuya meteórica carrera transcurrió entre 1795 y su muerte a la edad de 31 años en 1802. En los siete apretados años de profesión investigó no sólo la localización y patología de los órganos y estructuras, sino también su funcionamiento y contribución a los procesos de vida, muerte y enfermedad; publicó cinco tratados, siendo los más famosos las *Recherches physiologiques sur la vie et la mort* (1799) y la *Anatomie générale, appliquée à la physiologie et à la médecine* (1801). Como indica en el prefacio de sus *Recherches sur la vie et la mort*, se propuso unir el método experimental de Haller y Spallanzani con las perspectivas filosóficas de Bordeu.

"Vida", definía, "es el conjunto de funciones que se oponen a la muerte". De la vida se ocupa en la primera parte de las *Recherches*; de la muerte, en la segunda. Ninguna de las dos constituye un asunto fácil. La vida abarca dos vertientes: orgánica y animal. Las funciones de la vida orgánica, comunes a todos los seres vivos, animales y vegetales, son el nacimiento, la nutrición, el desarrollo y la muerte. Son internas al organismo. Las funciones de la vida animal comprenden la sensibilidad, percepción, volición y movilidad: ponen al individuo en relación con el mundo. Entre ambas, existe un hiato tajante: los órganos que prestan soporte a la vida animal —cerebro, oídos, nariz, boca, cuerdas vocales, músculos voluntarios y esqueleto— son bilateralmente simétricos; los órganos que realizan funciones de la vida orgánica —corazón, vasos, estómago, intestinos, vejiga, hígado— son de forma irregular o, en riñones y pulmones, desiguales.

En el centro del pensamiento de Bichat se encuentra la diferencia entre leyes vitales y leyes físicas. No cabe tender puentes entre ambas. Pero rechazó también reducir todos los fenómenos vitales a un solo principio, a la manera de Barthez. Para su doctrina Gillispie acuña el término *plurivitalismo*, en el sentido de lo propuesto por Bordeu: cada órgano posee su propia parte de la suma de las fuerzas vitales que animan al organismo. Magendie rechazaría la multiplicidad de tales fuerzas; reduciría, a su vez, los rasgos comunes del animal a dos tipos de operaciones: nutrición (el proceso en cuya virtud el cuerpo se compone y descompone) y movimiento (no sólo el locomotor, sino también la actividad del corazón, pulmones, sistema digestivo, etcétera). A su manera, Magendie aceptaba la idea de fuerza vital.

Hacia finales del siglo XVIII, Immanuel Kant imprime un nuevo giro en la concepción vitalista en su *Crítica del juicio teleológico*: los organismos se hallarían dotados de un propósito natural en el que todas las partes constituyen, de una forma armónica, fin y medios. El vitalismo gozó de predicamento hasta comienzos del siglo XX. Henri Bergson (1859-1941) y Hans Driesch (1867-1941) fueron vitalistas eximios y autores de cabecera de la juventud naturalista de Ernst Mayr. Driesch, un embriólogo de finura excepcional, supo describir el desarrollo anatómico con puntillosa precisión. Había comenzado su carrera en un campo mecanicista, el de la *Entwicklungsmechanik* (mecánica del desarrollo). Pero en su propio trabajo embriológico descubrió que un huevo entero de erizo de mar podía desarrollarse a partir de medio huevo, y eso le llevó a proponer una suerte de fuerza vital que supervisara el desarrollo, lo que él llamo "entelequia", algo especial y exclusivo de los seres vivos. En su tiempo era habitual indagar las causas químicas e incluso eléctricas de los fenómenos biológicos. Pensemos en Jacques Loeb, quien afirmaba haber mantenido un corazón latiendo en una solución durante 25 años; las así llamadas funciones vitales no eran más que funciones mecánicas.

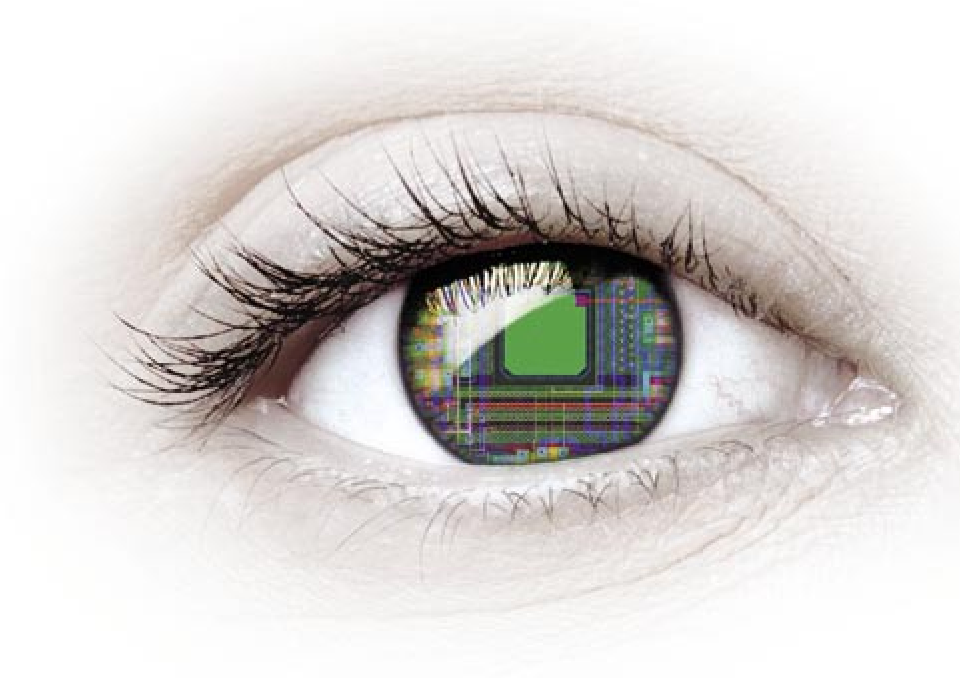
Ahora, ambas posturas parecen fenecidas. Nadie propone la tesis vitalista, ni parece necesario apelar al programa mecanicista de Loeb para contrarrestarlo. La vieja polémica se ha metamorfoseado en torno a la propia naturaleza de la biología. Frente al animal máquina de Descartes, se levanta la permanente cuestión de la singularidad de lo biológico, capaz de reproducirse, de abrirse al medio e interaccionar con él, de canalizar esa interacción en información. Los biólogos del desarrollo, en particular, plantean las cuestiones más punzantes: ¿puede acaso una máquina regenerar sus miembros amputados? ¿Puede una máquina replicarse a sí misma? ¿Pueden dos máquinas fundirse en una sola a la manera en que la fusión de dos gametos produce un cigoto? La lógica crítica de los vitalistas resultaba impecable, pero les faltó la prueba decisiva: dar con la *Lebenskraft* de Driesch.

Descartado el vitalismo, para matizar la resolución de la biología en la química y, por último, en la física, se debate ahora en torno al alcance de dicha reducción. En efecto, la relación entre las ciencias biológicas y las ciencias fisicoquímicas se ha denominado el problema del reduccionismo, que hunde sus raíces en el viejo apartado epistemológico, tan caro a los neopositivistas, de la reducción de teorías. Suelen distinguirse tres campos: ontológico, metodológico y epistemológico.

En el llamado reduccionismo ontológico, estructural o constitutivo, se supone que los organismos consisten en entidades fisicoquímicas, sin más; vale decir, combinaciones moleculares sin otras propiedades que las que dimanan de los átomos y las moléculas. El reduccionismo metodológico concierne al procedimiento científico que debe emplearse; esto es, si el enfoque adecuado es el estudio del nivel molecular, fisicoquímico, o si deben abordarse otros planos, (celular, orgánico, poblacional) en el caso de que se concediera a la biología un estatuto particular, irreductible. Por fin, el reduccionismo epistemológico, teórico o explicativo, atañe a la cuestión de si las teorías y leyes de la biología se deducen de las leyes y teorías de la física y la química; llevado a sus últimas consecuencias, el reduccionismo epistemológico sostiene que la biología terminará por desaparecer como ciencia, toda vez que representaría sólo un apartado de la física y la química. Este reduccionismo implacable no puede explicar las cuestiones fundamentales planteadas antes por la biología del desarrollo. Por lo que volvemos al principio. ¿Difiere de lo inerte lo vivo?

—LUIS ALONSO



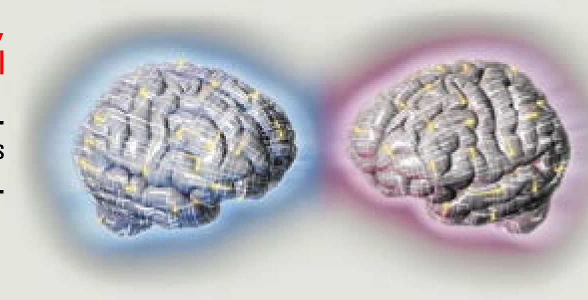


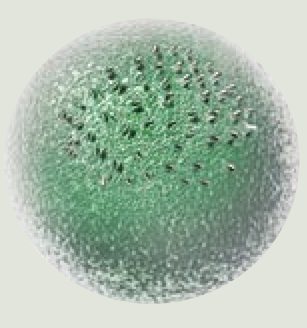
MICROCHIPS NEUROMORFICOS, por Kwabena Boahen

Mediante microcircuitos compactos que remedan la estructura neuronal del cerebro podrían fabricarse retinas artificiales de silicio que restituyan la visión, ojos robotizados y otros sensores inteligentes.

CEREBRO DE VARON, CEREBRO DE MUJER, por Larry Cahill

El cerebro del hombre y el de la mujer difieren en organización y actividad. La investigación de estas variaciones podría llevar a tratamientos específicos para cada sexo en trastornos como la depresión o la esquizofrenia.





¿SE PUEDE DETENER A LAS CLAMIDIAS?, por David M. Ojcius, Toni Darville y Patrik M. Bavoil

Las clamidias producen una enfermedad sexual muy extendida, son la principal causa de ceguera evitable y quizás intervengan en las afecciones cardiacas. Descubrimientos recientes sugieren nuevas vías para frenar su proliferación.

CITOQUINESIS EN CELULAS EUCARIOTAS, por Mercedes Pardo

En la citoquinesis, etapa final del ciclo celular, se forma una barrera que separará el material genético recién dividido en dos células hijas.

AGUJEROS NEGROS CUANTICOS, por Bernard J. Carr y Steve B. Giddings

Pronto podrían crearse agujeros negros en el laboratorio.

